APPENDIX A:

INTRODUCTION

Barry Glaz

Review Questions

1. Alpha and Beta rejected a H_o based on results with $p = 0.0498$ and did not reject when $p = 0.0501$. Were these good decisions?

a. Yes, we live and die by 0.05.

b. They should have rejected the H_o in both cases.

c. They should have accepted the H_o in both cases.

d. Had they also considered the effects of a Beta error, it is extremely likely that they would have either rejected or accepted the Ho in both cases.

Answer: a is technically correct in that we currently live and die by 0.05, but doing so is killing us. The best answer is d.

2. Rho did not understand statistics. You should ignore a significant interaction if at least one main effect is significant.

a. True

b. False

Answer: b. False. The grumpy ox had this one right.

3. Alpha and Beta's research on maximizing the morning snack of the oxen employees of Delta Oxlines should follow up with higher rates of coffee and donut.

a. True

b. False

Answer: a. True. Since oxcart-pulling distance increased at the highest rates of coffee and donut (1000 ml and 500 g, respectively), Alpha and Beta should conduct more research to find the rate at which the response to coffee and donut is maximized.

4. What is δαρν in English?

a. The equation of a complex mixed model.

b. A Type 5 error.

c. The ox fraternity in The Wondrous Land.

d. darn.

Answer: d. darn.

CHAPTER 1: ERRORS IN STATISTICAL DECISION MAKING

Kimberly Garland-Campbell

Software code

For hypothetical experiment, critical F values and beta values were calculated as below:

SAS

We created a dataset named ALPHA with two variables; the first variable is also named ALPHA and is a range of levels for Type 1 error from zero to 0.9999. The second variable is named FVALUE and is the *F* value associated with the effect that we would like to obtain (1.5, 4, and 9.285). The following code creates a new dataset named ERROR with the variables from the ALPHA dataset plus the following variables: PROB, NUMDF, DENDF, NONCENT, FCRIT, POWER, BETA, and AVEERROR, where PROB is the probability of alpha error, NUMDF is the numerator degrees of freedom for the effect, DENDF is the denominator degrees of freedom associated with the experimental error, NONCENT is the noncentrality parameter, FCRIT is the critical *F* value for tests of a significant difference in effects, POWER is the power of the test and BETA is 1-POWER or the Type 2 error associated with the test. AVEERROR is the average of the alphas and the beta errors for the various scenarios. The resulting ERROR dataset can be exported into a spreadsheet using Export Wizard in the File menu, or copied from the PROC PRINT statement. The ERROR dataset contains the data used in Table 1 and in Figures 1-3.

```
DATA ERROR;  SET ALPHA;
PROB=(1-ALPHA);
NUMDF=7;
DENDF=14;
NCP=NUMDF*FVALUE;
FCRIT=FINV(PROB, NUMDF, DENDF,0);
POWER=1-PROBF(FCRIT,NUMDF,DENDF,NCP);
BETA=1-POWER;
AVEERROR=(ALPHA+BETA)/2;
PROC PRINT DATA=ERROR;
RUN:
```

In R:

```
Error<-read.csv("alpha.csv")
Error$prob <- (1-(alpha))
Error$numdf<-7
Error$dendf<-14
Error<-transform(Error,ncp=(numdf*Fvalue))
Error<-transform(Error,fcrit=qf(alpha,numdf,dendf))
Error<-transform(Error,power=1-pf(fcrit,numdf,dendf,ncp))
Error<-transform(Error,beta=1-power)
Error<-transform(Error,aveterror=(alpha+beta)/2)
print(Error)
```

Yates Oat Experiment:

Data are available from the 'agridat' package in R as yates.oat.csv.

In this dataset, the x value is the columns and the y variable is the rows in the experimental design. The cultivar effect is named GEN and the manure effect is named NITRO.

```
SAS code using Proc Mixed:

PROC SORT DATA=YATES;
BY GEN NITRO BLOCK;
RUN;

PROC MIXED DATA=YATES METHOD=REML COVTEST PLOTS=ALL;
```

*This statement identifies the dataset, requests REML analysis and full residuals plots. The COVTEST statement requests tests of the covariance parameters;

```
TITLE 'YATES OAT SPLIT PLOT ANALYSIS';
```

* The title statement is an identifier and can be modified as needed;

```
CLASS GEN NITRO BLOCK;
```

*The class statement identifies the nitrogen and block variables as factors rather than nominal variables. In this analysis, nitrogen is considered to be a factor (categorical) variable, but it is conceivable that it would be a nominal (quantitative) variable in another type of analysis;

```
MODEL YIELD=GEN NITRO GEN*NITRO;
```

*The genotype and nitrogen treatments are fixed effects and replications are random so they are specified as such in the RANDOM statement;

```
RANDOM BLOCK BLOCK*GEN;
```

* The random statement specifies the two random effects for linear trend;

```
PROC MIXED DATA=YATES METHOD=REML COVTEST PLOTS=ALL;
TITLE 'YATES OAT SPLIT PLOT WITH LINEAR TREND';
CLASS GEN NITRO BLOCK X;
```

*The *X* variable is added to the classification statements to detect trends along the columns and is also added to the fixed effects below;

```
MODEL YIELD=GEN NITRO GEN*NITRO X;
RANDOM BLOCK BLOCK*GEN;
LSMEANS X GEN NITRO GEN*NITRO/ E CL;
RUN;
```

Similar models can be tested using R as described in the documentation for the agridat package: (agridat.pdf. pp. 318-319). A few additional pieces of code have been added to the code included in the agridat.pdf documentation for reasons described below. The code below is for the dataset available in the agridat package as well as for a new dataset containing the new randomization.

```
library(agridat)
require(lattice)
require(lme4)
require(lsmeans)
require(lucid)
```

#yield of oats in a split block experiment with four nitrogen levels, three oat cultivars, six blocks, total of 72 plots

import data from agridat package:

```
data("yates.oats")
dat<-yates.oats
```

#Plots the general layout of the data yields
#Includes the experimental design in the plots with the yield data.

```
desplot(yield ~ x*y, data=dat, out1=block, text=gen, col=nitro,
        cex=1, main="Yield Data for original Yates.oats Dataset")
```

#The plot shows that there is a linear gradient across the field.
Right-half of each block has lower yield.
Conduct split-plot analysis using lmer function in the 'lme4' package in R
#Have to note nitro as factor or the program will evaluate it with 1 df as a numerical variable

```
oatsp <- lmer(yield ~ factor(nitro) * gen + (1|block/gen), data=dat)
```

#print out all effects

```
summary(oatsp)
```

#print summary of fixed effects

```
anova(oatsp)
```

#print out fixed and random effects for model

```
fixef(oatsp)
ranef(oatsp)
```

#print out variance components for model

```
vc(oatsp)
```

#obtain lsmeans for the main effects and print them

```
lsmnit <- lsmeans(oatsp, "nitro")
lsmgen <- lsmeans(oatsp, "gen")
print(lsmnit)
print(lsmgen)
```

#plot residuals

```
qqmath(ranef(oatsp))
```

#Rerun model with a linear effect to correct for the linear trend.
Add a linear trend for column to the split-plot arrangement

```
oatsplin <- lmer(yield ~ x + factor(nitro) * gen + (1|block/gen),
data=dat)
summary(oatsplin)
anova(oatsplin)
fixef(oatsplin)
ranef(oatsplin)
```

#The residual variance is reduced

```
vc(oatsplin)
```

#The means do not change

```
lsmnitlin <- lsmeans(oatsplin, "nitro")
lsmgenlin <- lsmeans(oatsplin, "gen")
print(lsmnitlin)
print(lsmgenlin)
```

#The residuals plot has a better fit when the linear trend is accounted for

```
qqmath(ranef(oatsplin))
```

Review Questions, Answers:

True or False:

1. The central *F* distribution is calculated based on the numerator and error degrees of freedom.

TRUE. The *F* distribution depends on the ratio between the treatment or numerator and the error or denominator degrees of freedom.

2. Type 1 error should always be controlled below 5% whenever possible.

FALSE. While controlling α error below 0.05 or 5% is common, it is not always the most desirable option given the relative importance of α and β errors and the constraints to the experimental design.

3. The noncentrality parameter is associated with the effect size.

TRUE. The noncentrality parameter, λ, is determined by the effect size, the size of the experiment, and the unexplained experimental error.

4. Experiments should always be designed to obtain the minimum average error.

FALSE. This is a bit of a trick question. If money, space, time, and other similar constraints did not impact experimental design then this statement would be true, but because these all do impact our ability to conduct research, tradeoffs are required.

5. Effect sizes can be divided into those that measure differences between groups and those that measure association.

TRUE. While there are many statistics that can be used to measure effect size, this is a nice way to group them.

6. When spatial variation is discovered after the experiment is conducted, it will have to be included in the unexplained error for the experiment.

FALSE. Major trends can be detected and removed and other types of mixed model design that model covariance among experimental errors can be used to model spatial variation. But it is a good idea to anticipate spatial and temporal variation and block for it during the experimental design phase.

7. An experiment with a good deal of power will be associated with a lower probability of false positives.

FALSE. Actually, more power will have a lower probability of false negatives (β) and may or may not impact false positive error (α) depending on the experimental design.

8. The null hypothesis test is a valid approach to agronomic and environmental research.

TRUE. While criticized, the null hypothesis test still provides a good framework for decision making in agronomic and environmental research as long as the assumptions are met as described in this and other chapters in this book.

Solutions to Exercises:

Question 1:

In SAS:

```
PROC MIXED DATA=YATES METHOD=REML COVTEST PLOTS=ALL;
TITLE 'YATES OAT RCB ANALYSIS';
CLASS GEN NITRO BLOCK;
MODEL YIELD=GEN NITRO GEN*NITRO;
RANDOM BLOCK;
LSMEANS GEN NITRO GEN*NITRO/ E CL;
RUN;
```

In R:

```
oatrcb <- lmer(yield ~  factor(nitro) * gen + (1|block), data=dat)
summary(oatrcb)
```

```
anova(oatrcb)
fixef(oatrcb)
ranef(oatrcb)
vc(oatrcb)
```

#The means do not change

```
lsmnit <- lsmeans(oatrcb, "nitro")
lsmgen <- lsmeans(oatrcb, "gen")
print(lsmnit)
print(lsmgen)
qqmath(ranef(oatrcb))
```

Question 2:

In SAS:

```
PROC MIXED DATA=YATES METHOD=REML COVTEST PLOTS=ALL;
TITLE 'YATES OAT RCB ANALYSIS WITH LINEAR TREND';
CLASS BLOCK GEN NITRO X;
MODEL YIELD=X GEN NITRO GEN*NITRO;
RANDOM BLOCK;
LSMEANS X GEN NITRO GEN*NITRO/ E CL;
RUN;
```

In R:

```
oatrcblin <- lmer(yield ~ x + factor(nitro) * gen + (1|block/gen),
data=dat)
summary(oatrcblin)
anova(oatrcblin)
fixef(oatrcblin)
ranef(oatrcblin)
```

#The residual variance is reduced

```
vc(oatrcblin)
```

#The means do not change

```
lsmnitlin <- lsmeans(oatrcblin, "nitro")
lsmgenlin <- lsmeans(oatrcblin, "gen")
print(lsmnitlin)
print(lsmgenlin)
```

#The residuals plot has a better fit when the linear trend is accounted for

```
qqmath(ranef(oatrcblin))
```

QUESTION 3:

In SAS or R:

Same as above except replace *x* with *y*. (R is case sensitive so use lower case).

Question 4:

Notes:

Grand mean = 104.

25% of grand mean = 26.

Standard error of a difference can be calculated as $(2*MSE/24)^{1/2}$ for Geno; $(2*MSE/18)^{1/2}$ for Nitro and $(2*MSE/6)^{1/2}$ for their interaction. The *t* value for each of these can be calculated as 26/SED for each effect and the *F* value for the contrast as t^2. The following code can then be used to figure out experimental design parameters that work for each effect. An example for geno in the RCB analysis is below for alpha=0.05.

IN SAS:

```
DATA ERROR;  SET ALPHA;
TITLE 'ALPHA ERROR FOR GEN';
MSE=254.2; *MSE is from mixed model analysis of dataset;
```

```
SED=SQRT((2*MSE)/24);
ALPHA=0.05;
FVALUE=(26/SED)*(26/SED);
PROB=(1-ALPHA);
NUMDF=2;
DENDF=60;
NCP=NUMDF*FVALUE;
FCRIT=FINV(PROB, NUMDF, DENDF,0);
POWER=1-PROBF(FCRIT,NUMDF,DENDF,NCP);
BETA=1-POWER;
AVEERROR=(ALPHA+BETA)/2;
PROC PRINT DATA=ERROR;
RUN;
```

In R:

```
Error<-read.csv("alpha.csv")
Error$alpha<-0.05
Error$prob <- (1-(alpha))
Error$numdf<-2
Error$dendf<-60
Error$sed<-(sqrt(2*254/24))
Error$t<-(26/Error$sed)
Error$Fvalue<-(Error$t)*(Error$t);
Error<-transform(Error,ncp=(numdf*Fvalue))
Error<-transform(Error,fcrit=qf(alpha,numdf,dendf))
Error<-transform(Error,power=1-pf(fcrit,numdf,dendf,ncp))
Error<-transform(Error,beta=1-power)
Error<-transform(Error,aveterror=(alpha+beta)/2)
str(Error)
```

CHAPTER 2: ANALYSIS OF VARIANCE

Marla McIntosh

Supplement 1. Statbean Data.

Supplement 1. Statbean Data

Loc	Blk	Mulch	Ca_Trt	pH	Ca	Pct_Total	Trt_No
Central	1	0	0	5.58	765.604	5	1
Central	1	0	G1X	5.91	1072.395	5.93	2
Central	1	0	G2X	5.55	884.1182	13.33	3
Central	1	0	L1X	6.07	976.6795	7.86	4
Central	1	0	L2X	6.07	1030.769	12.86	5
Central	2	0	0	5.35	484.6031	29.23	1
Central	2	0	G1X	5.55	700.2342	22.5	2
Central	2	0	G2X	5.6	944.7611	24.44	3
Central	2	0	L1X	6.27	1651.486	15.38	4
Central	2	0	L2X	5.97	1159.154	24.29	5
Central	3	0	0	5.34	352.0018	28.7	1
Central	3	0	G1X	5.18	350.7162	34.55	2
Central	3	0	G2X	5.93	1041.485	25.83	3
Central	3	0	L1X	5.55	554.0077	22.61	4
Central	3	0	L2X	6.13	1397.836	18.33	5
West	1	0	0	5.89	1152.135	17.14	1
West	1	0	G1X	6.16	1730.585	10.71	2
West	1	0	G2X	5.66	1269.248	7.86	3
West	1	0	L1X	6.63	2309.693	25	4
West	1	0	L2X	6.52	1764.905	21.43	5
West	2	0	0	7.1	2673.196	28.89	1
West	2	0	G1X	6.89	2127.684	49.29	2
West	2	0	G2X	7.12	2806.581	40.83	3
West	2	0	L1X	7.24	2414.204	46.43	4
West	2	0	L2X	7.31	3146.425	23.57	5
West	3	0	0	6.76	2116.01	28.15	1
West	3	0	G1X	6.74	2522.883	22.96	2
West	3	0	G2X	6.77	1868.224	32.59	3
West	3	0	L1X	6.94	2513.69	30.37	4

West	3	0	L2X	6.97	2452.25	24.35	5
East	1	0	0	3.85	71.59624	0.77	1
East	1	0	G1X	3.93	84.90909	2.14	2
East	1	0	G2X	3.88	51.67797	0	3
East	1	0	L1X	4.12	94.65318	0.74	4
East	1	0	L2X	4.44	294.2328	2.96	5
East	2	0	0	4.21	153.2803	35.71	1
East	2	0	G1X	4.08	225.5879	16.43	2
East	2	0	G2X	4.18	242.5236	6.4	3
East	2	0	L1X	4.43	360.1202	38.57	4
East	2	0	L2X	4.76	465.0364	47.86	5
East	3	0	0	3.98	33.1985	0	1
East	3	0	G1X	4.09	79.34783	0	2
East	3	0	G2X	3.99	101.7181	0	3
East	3	0	L1X	3.92	22.76089	2.22	4
East	3	0	L2X	4.12	83.36722	1.6	5
Central	1	1	0	6.05	893.019	2.86	6
Central	1	1	G1X	5.81	930.887	4.29	7
Central	1	1	G2X	5.87	1203.349	9.29	8
Central	1	1	L1X	6.08	1084.533	6.67	9
Central	1	1	L2X	6.12	1450.82	9.29	10
Central	2	1	0	5.53	588.843	4.44	6
Central	2	1	G1X	5.69	823.8247	3.2	7
Central	2	1	G2X	5.57	860.6557	3.33	8
Central	2	1	L1X	5.7	863.8875	5.22	9
Central	2	1	L2X	5.78	817.8222	3.57	10
Central	3	1	0	5.47	646.8615	3.75	6
Central	3	1	G1X	6.09	1521.4	4.76	7
Central	3	1	G2X	5.08	548.4234	16.67	8
Central	3	1	L1X	6.08	1605.782	5.45	9
Central	3	1	L2X	6.29	1496.161	7.83	10
West	1	1	0	6.48	1799.348	21.54	6
West	1	1	G1X	6.42	1593.126	13.33	7
West	1	1	G2X	5.76	1171.701	14.62	8
West	1	1	L1X	6.58	2200.379	38.46	9
West	1	1	L2X	6.84	2782.864	24	10
West	2	1	0	7.02	2211.675	27.86	6
West	2	1	G1X	7.1	2497.6	51.43	7
West	2	1	G2X	7.15	2718.435	58.57	8
West	2	1	L1X	7.25	2562.135	61.43	9

West	2	1	L2X	7.28	2625.38	37.14	10
West	3	1	0	6.92	1977.072	40	6
West	3	1	G1X	6.67	1532.194	38.4	7
West	3	1	G2X	6.58	1954	19.05	8
West	3	1	L1X	6.93	2139.155	30	9
West	3	1	L2X	7.03	2470.402	41.6	10
East	1	1	0	4.03	12.2972	0	6
East	1	1	G1X	3.97	45.73493	0	7
East	1	1	G2X	3.92	97.56944	0	8
East	1	1	L1X	4.06	35.53188	1.48	9
East	1	1	L2X	4.12	56.58363	1.6	10
East	2	1	0	4.33	187.6108	15	6
East	2	1	G1X	4.21	305.4299	11.43	7
East	2	1	G2X	4.13	375.1605	17.14	8
East	2	1	L1X	4.67	416.7319	22.86	9
East	2	1	L2X	4.76	450	25	10
East	3	1	0	3.95	9.765314	0	6
East	3	1	G1X	3.91	173.9345	0	7
East	3	1	G2X	3.75	402.1609	1.6	8
East	3	1	L1X	4.21	123.3071	0.69	9
East	3	1	L2X	4.14	210.8767	1.48	10

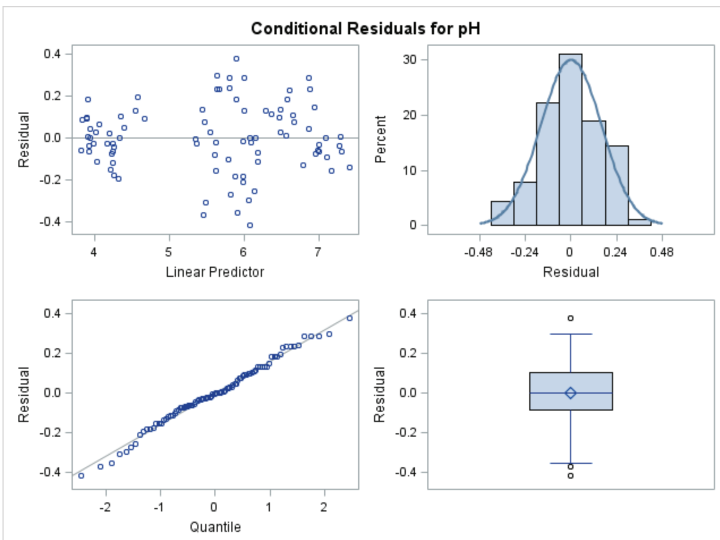
Supplement 2. SAS code - PROC MIXED for pH - by location (with and without contrast statements) and combined over locations.

Note: The data used for this program is provided as an Excel file 'Statbean Data.xlsx' in Supplement 1 and as a SAS dataset 'statbean.sas7bdat' in Supplement 6.

Replace the Data statement to use the dataset in Supplement 1.

```
data anova.statbean; set anova.statbean;
run;
proc sort; by loc;
run;
Title 'Statbean Data';
proc print;
run;
Title 'Mixed pH ANOVA by location without contrasts';
proc mixed data=anova.statbean plots=residualpanel method=type3; by
loc;
class Blk Mulch Ca_Trt;
model pH=Mulch Ca_Trt Mulch*Ca_Trt;
random Blk;
lsmeans Mulch Ca_Trt Mulch*Ca_Trt;
run;
Title 'Mixed pH ANOVA combined locations';
proc mixed data=anova.statbean plots=residualpanel method=type3;
class Loc Blk Mulch Ca_Trt;
model pH=Loc|Mulch|Ca_Trt;
random Blk(Loc);
```

```
lsmeans Loc|Mulch|Ca_Trt;
run;
Title 'Mixed pH ANOVA by location with contrasts';
proc mixed data=anova.statbean plots=residualpanel method=type3; by
loc;
class Blk Mulch Ca_Trt;
model pH=Mulch Ca_Trt Mulch*Ca_Trt;
random Blk;
lsmeans Mulch Ca_Trt Mulch*Ca_Trt;
contrast "Main Effect-Gypsum Linear" Ca_Trt -1 0 1 0 0/E;
contrast "Main Effect-Gypsum Quadratic" Ca_Trt 1 -2 1 0 0/E;
contrast "Main Effect-Lime Linear" Ca_Trt -1 0 0 0 1/E;
contrast "Main Effect-Lime Quadratic" Ca_Trt 1 0 0 -2 1/E;
contrast "Main Effect-Lime vs Gypsum" Ca_Trt 0 1 1 -1 -1/E;
contrast "Interaction-Gypsum Linear*Mulch" Ca_Trt*Mulch -1 0 1 0 0 1 0 -1 0 0/E;
contrast "Interaction-Gypsum Quadratic* Mulch" Ca_Trt*Mulch 1 -2 1 0 0 -1 2 -1 0 0/E;
contrast "Interaction-Lime Linear*Mulch" Ca_Trt*Mulch -1 0 0 0 1 1 0 0 0 -1/E;
contrast "Interaction-Lime Quadratic*Mulch" Ca_Trt*Mulch 1 0 0 -2 1 -1 0 0 2 -1/E;
contrast "Interaction-Lime vs Gypsum*Mulch" Ca_Trt*Mulch 0 1 1 -1 -1 0 -1 -1 1 1/E;
run;
```



Supplement 3. Provided in the electronic supplemental materials.

Supplement 4. SAS code for ANOVA for pH, Ca, and Pct_Total

```
data anova.statbean; set anova.statbean;
proc sort; by loc;
run;
Title 'Statbean Data';
proc print;
run;
Title 'Mixed pH ANOVA by location';
proc mixed data=anova.statbean plots=residualpanel method=type3; by
loc;
class Blk Mulch Ca_Trt;
model pH=Mulch Ca_Trt Mulch*Ca_Trt;
random Blk;
lsmeans Mulch Ca_Trt Mulch*Ca_Trt;
run;
Title 'Mixed pH ANOVA combined locations';
proc mixed data=anova.statbean plots=residualpanel method=type3;
```

```
class Loc Blk Mulch Ca_Trt;
model pH=Loc|Mulch|Ca_Trt;
random Blk(Loc);
lsmeans Loc|Mulch|Ca_Trt;
run;
Title 'Mixed Calcium ANOVA by location';
proc mixed data=anova.statbean plots=residualpanel method=type3; by
loc;
class Blk Mulch Ca_Trt;
model Ca=Mulch Ca_Trt Mulch*Ca_Trt;
random Blk;
lsmeans Mulch Ca_Trt Mulch*Ca_Trt;
run;
Title 'Mixed Calcium ANOVA combined locations';
proc mixed data=anova.statbean plots=residualpanel method=type3;
class Loc Blk Mulch Ca_Trt;
model Ca=Loc|Mulch|Ca_Trt;
random Blk(Loc);
lsmeans Loc|Mulch|Ca_Trt;
run;
Title 'Mixed Percent total ANOVA by location';
proc mixed data=anova.statbean plots=residualpanel method=type3; by
loc;
class Blk Mulch Ca_Trt;
model Pct_total=Mulch Ca_Trt Mulch*Ca_Trt;
random Blk;
lsmeans Mulch Ca_Trt Mulch*Ca_Trt;
run;
Title 'Mixed Percent total ANOVA combined locations';
proc mixed data=anova.statbean plots=residualpanel method=type3;
class Loc Blk Mulch Ca_Trt;
model Pct_total=Loc|Mulch|Ca_Trt;
random Blk(Loc);
lsmeans Loc|Mulch|Ca_Trt;
run;
```

Supplement 5 provided in the electronic supplemental materials.

Answers to Review Questions

1. True
2. False
3. False. The F-value numerator is the treatment MS which includes both Var(Treatment) and Var(Residual). See the EMS.
4. False, the Type 1 error rate is a probability set by the researcher.
5. True

CHAPTER 3: BLOCKING PRINCIPLES FOR BIOLOGICAL EXPERIMENTS

Michael D. Casler

Example 1. Conduct a linear mixed model ANOVA from an augmented design.

Problem: Augmented designs are unbalanced, specifically with reference to test treatments that are typically unreplicated.

Solution: Residual or error variances must be estimated from replicated treatments, which should be arranged in a manner that also allows estimation and removal of some spatial variation within the experimental area. Estimates for unreplicated treatments are then adjusted for spatial variation.

Example: Twenty-one soybean cultivars were evaluated in an augmented design, with four cultivars arranged in three randomized complete blocks and the other 17 cultivars each represented in only one of the three blocks (Scott and Milliken, 1993). Each column of the data set below represents one block.

SAS Code: The following code gives an ANOVA with a separate F test for check cultivars and test cultivars. It also provides adjusted cultivar means and standard errors for the four check cultivars and the 17 test cultivars. Note that we are using a trick that will allow SAS Proc Mixed to compute a separate p value for check cultivars, which are replicated, and for test cultivars, which are not replicated. The trick is to recode the cultivar number into two sets of numbers. The first set, c, codes the checks and has $c = 0$ for all the test cultivars. The second set, x, codes the test cultivars and has $x = 0$ for the check cultivars. The ANOVA model is then set up with two terms: check cultivars and test cultivars nested within check cultivars.

```
options nocenter;
data a; input entryno entry$ y1 y2 y3; datalines;
1    Sibley     4098  4060  4283
1    Sibley        .     .  3952
2    Hardin     4020  4414  3571
3    Weber      4440  3835  4154
4    Kato       3860  3865  3674
5    TEgg       2169     .     .
6    Harlon     3250     .     .
7    Rampage    3807     .     .
8    Steele     4068     .     .
9    Vinton     3871     .     .
10   Vinton81   3838     .     .
11   BSR101        .  4244     .
12   Norsoy        .  3290     .
13   WBlack        .  3019     .
14   Mandarin      .  3506     .
15   Hark          .  4384     .
16   Coles         .  4148     .
17   Hodgson78     .     .  4167
18   Lakota        .     .  4023
19   Mandan507     .     .  2435
20   Bert          .     .  4595
```

```
21   Leslie       .     .   3957
;
data b; set a;
yield=y1; block=1; output;
yield=y2; block=2; output;
yield=y3; block=3; output;
drop y1-y3; run;
data c; set b;
x=entryno; if x<5 then x=0;
c=entryno; if c>4 then c=0;
proc mixed; class block x c;
model yield = c x(c);
random block;
lsmeans x(c);
run;
```

SAS Output: The output below contains the mixed models ANOVA and the least squares means for all 21 cultivars.

```
The Mixed Procedure

                  Model Information
Data Set                     WORK.C
Dependent Variable           yield
Covariance Structure         Variance Components
Estimation Method            REML
Residual Variance Method     Profile
Fixed Effects SE Method     Model-Based
Degrees of Freedom Method    Containment

             Class Level Information
Class    Levels    Values
block         3    1 2 3
x            18    0 5 6 7 8 9 10 11 12 13 14 15
                   16 17 18 19 20 21
c             5    0 1 2 3 4

            Dimensions
Covariance Parameters            2
Columns in X                    27
Columns in Z                     3
Subjects                         1
Max Obs per Subject             30
          Number of Observations
Number of Observations Read             66
Number of Observations Used             30
Number of Observations Not Used         36

                      Iteration History
Iteration    Evaluations    -2 Res Log Like       Criterion
        0              1      130.48103794
        1              1      130.48103794      0.0000000

                    Convergence criteria met.

The Mixed Procedure
Covariance Parameter

      Estimates
Cov Parm     Estimate
block               0
Residual        68856
           Fit Statistics
-2 Res Log Likelihood          130.5
AIC (Smaller is Better)         132.5
AICC (Smaller is Better)        133.1
```

```
BIC (Smaller is Better)          131.6
        Type 3 Tests of Fixed Effects
              Num      Den
Effect         DF       DF    F Value    Pr > F
c               4        7       3.61    0.0667
x(c)           16        7       6.50    0.0090
                           Least Squares Means
                                  Standard
Effect    x     c   Estimate      Error      DF    t Value     Pr > |t|
x(c)      5    0    2169.00       262.41     7       8.27      < 0.0001
x(c)      6    0     3250.00      262.41     7      12.39      < 0.0001
x(c)      7    0     3807.00      262.41     7      14.51      < 0.0001
x(c)      8    0     4068.00      262.41     7      15.50      < 0.0001
x(c)      9    0     3871.00      262.41     7      14.75      < 0.0001
x(c)     10    0     3838.00      262.41     7      14.63      < 0.0001
x(c)     11    0     4244.00      262.41     7      16.17      < 0.0001
x(c)     12    0     3290.00      262.41     7      12.54      < 0.0001
x(c)     13    0     3019.00      262.41     7      11.51      < 0.0001
x(c)     14    0     3506.00      262.41     7      13.36      < 0.0001
x(c)     15    0     4384.00      262.41     7      16.71      < 0.0001
x(c)     16    0     4148.00      262.41     7      15.81      < 0.0001
x(c)     17    0     4167.00      262.41     7      15.88      < 0.0001
x(c)     18    0     4023.00      262.41     7      15.33      < 0.0001
x(c)     19    0     2435.00      262.41     7       9.28      < 0.0001
x(c)     20    0     4595.00      262.41     7      17.51      < 0.0001
x(c)      0    1     4098.25      131.20     7      31.24      < 0.0001
x(c)      0    2     4001.67      151.50     7      26.41      < 0.0001
x(c)      0    3     4143.00      151.50     7      27.35      < 0.0001
x(c)      0    4     3799.67      151.50     7      25.08      < 0.0001
```

Results and Conclusions: Check cultivars differed from each other with a p value of 0.07, while the test cultivars had $p < 0.01$. The check cultivars have 4 df in the numerator because this term is testing differences among five means: Sibley, Hardin, Weber, Cato, and the mean of all 17 test cultivars. Least squares means allow the researcher to choose the best test cultivars for further, more advanced, testing. Least squares means are adjusted for block effects, but not for spatial variation on a finer scale. Note that there are three standard error values, each one corresponding to $r = 1$, $r = 4$, or $r = 5$ experimental units per cultivar (262, 151, and 131, respectively).

Example 2. Make a logical and objective decision regarding whether or not random design effects should be retained in the final model for publication purposes.

Problem: Modern mixed models analysis is often taught in a manner that encourages researchers to used reduced models, containing only those terms that are important. This practice results in pooling random design effects with residual effects. There are three philosophies that can be employed in pooling when it is clear that a design component is small or nonsignificant: always pool, never pool, or pool using an objective decision tool that seeks to avoid Type 2 errors.

Solution: The example below will illustrate how to employ a likelihood ratio test to quantify a p-value for a random design component, then set a decision rule for "calling" that term significant or nonsignificant, with the final result to either include or exclude that term form the model. The methodology is based on the concepts and philosophy of Carmer et al. (1969) but using modern likelihood ratio tests, rather than *F* tests.

Example: The data are percentage survivorship of 14 Italian ryegrass (*Lolium multiflorum* Lam.) cultivars planted in factorial combination with three seeding rates (200, 400, and 800 seeds m^{-2}). The experiment was designed as a randomized complete block with four replicates and randomized with seeding rates as whole plots and cultivars as subplots.

SAS Code: There are two blocks of data in the SAS code below. The first consists of the field map in 24 rows × 7 columns: rows are identified by row number, rep number, and seeding rate, while the data in each column are the cultivar numbers (1 through 14). The second block of data consists of percentage survivorship for the 24 × 7 grid. The two blocks of data are merged together by row number, prior to conducting the analyses of variance.

```
options nocenter;
data a; input row rep rate x1-x7; datalines;
 1      4      40      1      11      13      9      4      7      10
 2      4      40      5       8       2      6      3     12      14
 3      4      80      3      11       6      1     10      7       8
 4      4      80     14       9       4     12      5      2      13
 5      4      20      4       8       2     11      1     10      12
 6      4      20      3       9      14      7      6     13       5
 7      3      80      4       5       2      3     10      6       1
 8      3      80      8      12       7     14      9     11      13
 9      3      40      5       7       9      8      4     14      11
10      3      40      1       2       3     10     13     12       6
11      3      20     11       5       4      3     13      7       1
12      3      20      6      14       9      8      2     10      12
13      2      40     13       9       2     12      6      8       5
14      2      40     11      14      10      3      4      1       7
15      2      20     12      13       7      1     11     10       4
16      2      20      6      14       5      9      3      8       2
17      2      80     14       1       9     12      2      3       6
18      2      80      4      11       8      5      7     10      13
19      1      80      5       2       8      9      3     14      11
20      1      80     12       1      13     10      7      4       6
21      1      20     11       1       8     13     14      7      10
22      1      20      5       3       6      4      9      2      12
23      1      40     10       1      12     11      4      8       2
24      1      40      6       5      14     13      3      7       9
;
data aa; set a;
cultivar=x1; col=1; output;
cultivar=x2; col=2; output;
cultivar=x3; col=3; output;
cultivar=x4; col=4; output;
cultivar=x5; col=5; output;
cultivar=x6; col=6; output;
cultivar=x7; col=7; output;
drop x1-x7;
proc sort; by row col; run;
data b; input row x1-x7; datalines;
 1     50      5      5      5      5      5      5
 2      5     50     40      5      5      5     10
 3      5      5      5     40      5     10     25
 4     50      5     60      5     20     45      5
 5      5     20      5      5     20      5      5
 6      5      5     10      5      5      5      5
 7     20      5     70      5      5     10     50
 8     40      5      5     60     10     15      5
 9      5     10      5     80     50     50      5
```

```
10      30      80       5       5      10       5       5
11       5       5      20       5       5      10      40
12       5       5       5      40      30       5       5
13       5       5      40       5       5      30       5
14       5      15       5       5      50      30      20
15       5       5       5      15       5       5      35
16       5      15       5       5       5      25      50
17      20      75      10       5      60       5       5
18      30       5      60      10      30       5       5
19       5      80      75       5       5      30      15
20       5      60      15       5      40      55      10
21       5      10      20       5       5       5       5
22       5       5       5      10       5      40       5
23       5      25       5       5      40      30      65
24       5       5      40       5       5       5       5
;
data bb; set b;
gc95=x1; col=1; output;
gc95=x2; col=2; output;
gc95=x3; col=3; output;
gc95=x4; col=4; output;
gc95=x5; col=5; output;
gc95=x6; col=6; output;
gc95=x7; col=7; output;
drop x1-x7;
proc sort; by row col; run;
data c; merge aa bb; by row col;
libname arfs '~/arfs';
proc mixed cl; class rep rate cultivar;
model gc95 = rate|cultivar;
random rep rep*rate;
lsmeans cultivar;
run;
proc mixed cl; class rep rate cultivar;
model gc95 = rate|cultivar;
random rep;
lsmeans cultivar;
run;
```

SAS Output: The output consists of mixed models analysis results and least squares means for cultivars for two different models. According to the model, least squares means for seeding rates and the cultivar × seeding rate interaction would also be important output, but have been left out here for brevity. The first model is the full model that includes three random effects: Blocks, Error(a), and Error(b) of the split-plot randomization. The second model is based on the visual observation that Error(a) is very small any may not actually be significant. The second model is identical to the first, except that it excludes Error(a), collapsing that term into Error(b), effectively treating this analysis as a simple randomized complete block without the split-plot randomization restriction.

```
The Mixed Procedure

                  Model Information
Data Set                     WORK.C
Dependent Variable           gc95
Covariance Structure         Variance Components
Estimation Method            REML
Residual Variance Method     Profile
Fixed Effects SE Method      Model-Based
Degrees of Freedom Method    Containment

              Class Level Information
```

```
Class       Levels    Values
rep              4    1 2 3 4
rate             3    20 40 80
cultivar        14    1 2 3 4 5 6 7 8 9 10 11 12 13
                      14

            Dimensions
Covariance Parameters           3
Columns in X                   60
Columns in Z                   16
Subjects                        1
Max Obs per Subject           168

          Number of Observations
Number of Observations Read            168
Number of Observations Used            168
Number of Observations Not Used          0

                     Iteration History
Iteration    Evaluations    -2 Res Log Like       Criterion
        0              1      1013.63658070
        1              1      1011.54388284      0.00000000
                   Convergence criteria met.

            Covariance Parameter Estimates
Cov Parm    Estimate   Alpha    Lower    Upper
rep          2.2003    0.05    0.2619   1.377E10
rep*rate     4.3932    0.05    0.7764   33250 <--- This is Error(a)
Residual     108.39    0.05   85.2069    142.56

           Fit Statistics
-2 Res Log Likelihood         1011.5
AIC (Smaller is Better)        1017.5
AICC (Smaller is Better)       1017.7
BIC (Smaller is Better)        1015.7

          Type 3 Tests of Fixed Effects
                  Num     Den
Effect            DF      DF    F Value    Pr > F
rate                2       6      13.53    0.0060
cultivar           13     117      29.20    < 0.0001
rate*cultivar      26     117       1.96    0.0082

                             Least Squares Means
                                     Standard
Effect     cultivar    Estimate      Error     DF    t Value    Pr > |t|
cultivar      1       37.0833     3.1541     117     11.76    < 0.0001
cultivar      2       50.4167     3.1541     117     15.98    < 0.0001
cultivar      3        5.0000     3.1541     117     1.59      0.115
cultivar      4       31.6667     3.1541     117     10.04    < 0.0001
cultivar      5       6.6667     3.1541      117      2.11    0.0367
cultivar      6       5.8333     3.1541      117      1.85   0.0669
cultivar      7       12.5000     3.1541     117      3.96    0.0001
cultivar      8       41.2500     3.1541     117     13.08    < 0.0001
cultivar      9       5.8333      3.1541     117      1.85   0.0669
cultivar     10       5.0000      3.1541     117      1.59    0.1156
cultivar     11       6.6667      3.1541     117      2.11    0.0367
cultivar     12       5.0000      3.1541     117      1.59    0.1156
cultivar     13       6.2500      3.1541     117      1.98   0.0499
cultivar     14       25.8333      3.1541     117      8.19    < 0.0001

The Mixed Procedure
                  Model Information
Data Set                       WORK.C
Dependent Variable             gc95
Covariance Structure           Variance Components
Estimation Method              REML
```

```
Residual Variance Method     Profile
Fixed Effects SE Method      Model-Based
Degrees of Freedom Method    Containment

             Class Level Information
Class       Levels    Values
rep              4    1 2 3 4
rate             3    20 40 80
cultivar        14    1 2 3 4 5 6 7 8 9 10 11 12 13
                      14

           Dimensions
Covariance Parameters            2
Columns in X                    60
Columns in Z                     4
Subjects                         1
Max Obs per Subject            168

          Number of Observations

Number of Observations Read            168
Number of Observations Used            168
Number of Observations Not Used          0

                      Iteration History
Iteration    Evaluations    -2 Res Log Like       Criterion
        0              1      1013.63658070
        1              1      1012.20561338      0.00000000
                    Convergence criteria met.

           Covariance Parameter Estimates
Cov Parm     Estimate     Alpha       Lower       Upper
rep            3.5932      0.05      0.7115     3947.53
Residual       111.39      0.05     88.0562      145.45

          Fit Statistics
-2 Res Log Likelihood          1012.2
AIC (Smaller is Better)        1016.2
AICC (Smaller is Better)       1016.3
BIC (Smaller is Better)        1015.0

          Type 3 Tests of Fixed Effects
                  Num     Den
Effect             DF      DF    F Value    Pr > F
rate                2     123      20.64    < 0.0001
cultivar           13     123      28.41    < 0.0001
rate*cultivar      26     123       1.91    0.0103

                            Least Squares Means
                                  Standard
Effect     cultivar    Estimate     Error     DF    t Value    Pr > |t|
cultivar    1         37.0833      3.1907    123    11.62      < 0.0001
cultivar    2         50.4167      3.1907    123    15.80      < 0.0001
cultivar    3          5.0000      3.1907    123     1.57      0.1197
cultivar    4         31.6667      3.1907    123     9.92      < 0.0001
cultivar    5          6.6667      3.1907    123     2.09      0.0387
cultivar    6          5.8333      3.1907    123     1.83     0.0699
cultivar    7         12.5000      3.1907    123     3.92      0.0001
cultivar    8         41.2500      3.1907    123    12.93      < 0.0001
cultivar    9          5.8333      3.1907    123     1.83     0.0699
cultivar   10          5.0000      3.1907    123     1.57      0.1197
cultivar   11          6.6667      3.1907    123     2.09      0.0387
cultivar   12          5.0000      3.1907    123     1.57      0.1197
cultivar   13          6.2500      3.1907    123     1.96      0.0524
cultivar   14         25.8333      3.1907    123     8.10      < 0.0001
```

Results and Conclusions: The first step in interpreting the results is to conduct a likelihood ratio test of Error(a), in other words, testing the null hypothesis Ho:σ^2_a=0, where σ^2_a is the Error(a) covariance component (estimated as 4.39 in the first mixed models ANOVA). The formula for this test is to compute the difference between the −2RLL (residual log likelihood) values and divide them by the difference in the number of covariance parameters. This becomes: (1012.2 − 1011.5)/(3 − 2) = 0.7. This value is tested as a χ^2 variate with 3 − 2 = 1 df, resulting in p = 0.40. Using the Carmer et al. (1969) guideline, that the null hypothesis should not be rejected unless p < 0.50, we would *not* reject this null hypothesis. In other words, we would *not* reduce the model; we would use the first mixed models analysis above as our final result, assuming all other diagnostics and results are correct. Note the change in denominator df between the two analyses, from 117 to 123—this is correctly indicating that the second estimate of residual variance is the pooled Error(a) and Error(b) value with 117 + 6 = 123 df. Also, note that the standard error of a cultivar mean actually went up after pooling, from 3.15 to 3.19—this is another sign that pooling, or model reduction, is a bad idea in this case.

Example 3. Review Exercise. Arranging blocks and blocking patterns for future experiments.

Problem: Decisions on exactly how to arrange blocks in many field experiments are very difficult, because there is often little information available to determine if there are gradients and the patterns of any gradients. This is especially true on agricultural experiment stations, which are often located on sites that have a uniform visual appearance.

Solution: Conduct retrospective analyses of previous experiments on a given site to determine the size, scale, and direction of spatial variation.

Example: Step 1. Consider a randomized complete block experiment that has been completed on a given site. In this example, the experiment was designed with 12 rows and 6 columns. Conduct an ANOVA on the data and output the residuals from the ANOVA. The blocking pattern of the previous experiment is not important, because the residuals are adjusted for both block and treatment effects, representing only pure spatial variation present in the field, at the scale of the experimental unit. In this case, we added the grand mean back to each residual, so that the data appears in the original units of measurement (Mg ha^{-1} of plant biomass). Step 2. Create dummy variables that can be used to simulate different blocking patterns on this site. Use each dummy variable to conduct a simulated one-way ANOVA with two sources of variation: among dummy blocks and within dummy blocks. Choose a design pattern with a high F ratio for dummy blocks (low within-block variance).

SAS Code: The following data are residuals + grand mean from a field experiment conducted in 6 columns × 12 rows, so that treatment and block variation has already been removed. The SAS code creates several combinations of rows and columns, after which it computes a one-way ANOVA for 22 different blocking arrangements and sizes. Comparison of the residual variances across the different blocking arrangements and sizes can help a researcher better understand the scale and dimensions of spatial variation within a field and plan better blocking schemes for future experiments.

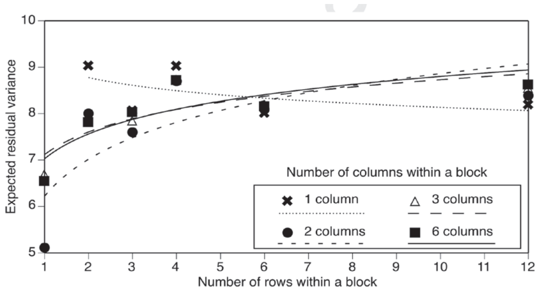
```
options nocenter;
data a; input row y1 y2 y3 y4 y5 y6;
datalines;
1         7.85      7.86      9.13      9.63     10.52     10.24
2        10.08     10.78     15.18     11.66     12.57     13.60
3        14.41     10.19     16.11     13.53     12.66     14.37
4        14.18      9.77     13.85     12.99     13.09     13.38
5         9.87      6.24      7.27      9.68     13.91      9.62
6        11.63     14.00     15.34     10.99     13.04     13.21
7        11.39     12.33     18.57     18.65     12.79     12.10
8        19.20     13.75     11.47      9.43     17.94      9.26
9        19.03     13.59     11.52     14.09     16.19     12.21
10       11.31      5.52     11.40      9.71     13.51     10.65
11        8.23     13.27     14.51     13.16     17.15     16.81
12       12.59     11.27     11.44     13.58     15.69     13.87
;
data b; set a;
yield=y1; col=1; output;
yield=y2; col=2; output;
yield=y3; col=3; output;
yield=y4; col=4; output;
yield=y5; col=5; output;
yield=y6; col=6; output;
drop y1-y6; run;
data c; set b;
r1=row;
r2=int((row+1)/2);
r3=int((row+2)/3);
r4=int((row+3)/4);
r6=int((row+5)/6);
c1=col;
c2=int((col+1)/2);
c3=int((col+2)/3);
proc glm; class r1 c1; model yield = r1*c1;
proc mixed; class r2 c1; model yield = r2*c1;
proc mixed; class r3 c1; model yield = r3*c1;
proc mixed; class r4 c1; model yield = r4*c1;
proc mixed; class r6 c1; model yield = r6*c1;
proc mixed; class c1; model yield = c1;
proc mixed; class r1 c2; model yield = r1*c2;
proc mixed; class r2 c2; model yield = r2*c2;
proc mixed; class r3 c2; model yield = r3*c2;
proc mixed; class r4 c2; model yield = r4*c2;
proc mixed; class r6 c2; model yield = r6*c2;
proc mixed; class c2; model yield = c2;
proc mixed; class r1 c3; model yield = r1*c3;
proc mixed; class r2 c3; model yield = r2*c3;
proc mixed; class r3 c3; model yield = r3*c3;
proc mixed; class r4 c3; model yield = r4*c3;
proc mixed; class r6 c3; model yield = r6*c3;
proc mixed; class c3; model yield = c3;
proc mixed; class r1; model yield = r1;
proc mixed; class r2; model yield = r2;
proc mixed; class r3; model yield = r3;
proc mixed; class r4; model yield = r4;
proc mixed; class r6; model yield = r6; run;
```

SAS Output: The output from this SAS code results in one GLM ANOVA, which provides the variance among the raw yield values. This value of 8.63 is the estimated residual variance expected from the use of a completely randomized design on this site. The remaining output consists of the results from 22 mixed models ANOVAs, only the first of which is shown below. The critical item here is the Residual Covariance Parameter estimate of 9.0313. The remaining output shown below consists of the "Residual" line from all 22 Proc Mixed ANOVAs. Each of the 22 Proc Mixed ANOVAs provides a residual variance that would be expected for a different blocking design. For example, the first one corresponds to blocks that contain two experimental units in an arrangement of 2 rows × 1 column, while the last one corresponds to blocks that contain 36 experimental units in an arrangement of 6 rows × 6 columns.

```
The GLM Procedure
              Class Level Information

Class         Levels    Values

r1                12    1 2 3 4 5 6 7 8 9 10 11 12
c1                 6    1 2 3 4 5 6

Number of Observations Read          72
Number of Observations Used          72

The GLM Procedure
Dependent Variable: yield

                              Sum of
Source             DF        Squares     Mean Square    F Value    Pr > F

Model              71    612.7471875       8.6302421          .         .
Error               0      0.0000000        .
Corrected Total    71    612.7471875

R-Square     Coeff Var      Root MSE    yield Mean
1.000000          .              .        12.49458

Source             DF      Type I SS     Mean Square    F Value    Pr > F
r1*c1              71    612.7471875       8.6302421          .         .

The Mixed Procedure
                         Model Information
Data Set                     WORK.C
Dependent Variable           yield
Covariance Structure         Diagonal
Estimation Method            REML
Residual Variance Method     Profile
Fixed Effects SE Method     Model-Based
Degrees of Freedom Method    Residual

             Class Level Information
Class    Levels    Values
r2           6    1 2 3 4 5 6
c1           6    1 2 3 4 5 6

            Dimensions
Covariance Parameters            1
Columns in X                    37
Columns in Z                     0
Subjects                         1
Max Obs per Subject             72

          Number of Observations
Number of Observations Read              72
Number of Observations Used              72
Number of Observations Not Used           0

Covariance Parameter
      Estimates
Cov Parm     Estimate
Residual       9.0313
            Fit Statistics
-2 Res Log Likelihood          206.3
```

```
AIC (Smaller is Better)          208.3
AICC (Smaller is Better)         208.5
BIC (Smaller is Better)          209.9

        Type 3 Tests of Fixed Effects
              Num      Den
Effect        DF       DF     F Value     Pr > F
r2*c1          35       36       0.91     0.6093

Residual       9.0313
Residual       8.0655
Residual       9.0267
Residual       8.0114
Residual       8.1919
Residual       5.1116
Residual       8.0005
Residual       7.5977
Residual       8.6949
Residual       8.0895
Residual       8.3881
Residual       6.6952
Residual       7.89
Residual       7.8557
Residual       8.7369
Residual       8.1976
Residual       8.5585
Residual       6.5478
Residual       7.8181
Residual       8.0338
Residual       8.7223
Residual       8.1588
```

Results and Conclusions: The best way to visualize these results is to organize them into a graph. The residual variances above were matched up with two additional columns of data, the number of rows within the blocks and the number of columns within the blocks. The graph below illustrates the relationships. Clearly, the best three blocking scenarios were based on a single row across 2, 3, or 6 columns. All scenarios with multiple rows were relatively inefficient, indicating that most of the spatial variation was in the direction of rows. These results indicate that future experiments in this field should make all attempts to block out the variability associated with rows.



CHAPTER 4: POWER AND REPLICATION—DESIGNING POWERFUL EXPERIMENTS

Michael D. Casler

EXERCISE #1.

Predict the power of a future hypothetical experiment using the probability distribution method.

Solution

Using estimates of residual variances and covariance parameter estimates of other random factors estimated from previous experiments, it is possible to predict the power of future experiments under a wide range of design scenarios. From these predictions, it is then possible to make an intelligent assessment and comparison of different designs and choose an optimal design that balances statistical power with financial cost.

Example

Consider a proposed completely randomized design in which treatments are replicated as shown in Figure 1C and, additionally, multiple sampling (observational) units are created within each experimental unit, from which one data point is collected on each observational unit. This could equally apply to experiments in the field, glasshouse, laboratory, or benchtop. Replicates are nested within treatments and sampling units are nested within experimental units. The goal is to detect a difference between treatment means of 5% at a Type 1 error rate of $\alpha = 0.05$. Prior estimates of experimental error (5) and sampling error (10) are available (covariance parameter estimates of random effects from previous experiments); these values are defined in the "parms" statement.

SAS code

The SAS code below computes the expected power for the design described above with $r = 4$ replicates per treatment and $s = 2$ sampling units per experimental unit. It first creates the representative data set with treatment means of 95 and 100, following which Proc GLIMMIX is used to compute the F-ratio within the representative data set. The final set of computations generates the non-centrality parameter, assuming a normal distribution, which then leads to the power computation. Any of the input parameters can be changed to allow investigation and comparison of different designs.

```
options nocenter;
data a; input trt y;
*CREATE A REPRESENTATIVE DATA SET WITH 4 REPS AND 2 SAMPLES PER EXPTL
UNIT;
do rep=1 to 4 by 1;
do samples=1 to 2 by 1;
```

```
output;
end;
end;
*CREATE TREATMENT MEANS WITH THE DESIRED DETECTION VALUE;
datalines;
1 95
2 100
run;
*COMPUTE THE NON-CENTRALITY PARAMETER;
data b; set a;
proc glimmix; class trt rep;
model y = trt;
random rep(trt);
*INPUT PARAMETER ESTIMATES FOR EXPTL AND SAMPLING ERRORS;
parms (5)(10) / hold=1,2;
ods output tests3=power_terms;
*COMPUTE POWER OF THE TEST;
data power; set power_terms;
alpha=0.05;
ncparm=numdf*Fvalue;
F_critical=finv(1-alpha, numdf, dendf, 0);
power=1-probf(F_critical, numdf, dendf, ncparm);
proc print;
run;
```

SAS output

The output consists of one run of Proc GLIMMIX, including all relevant diagnostic and estimation information. The last line is the result of the "proc print" statement, printing out the results of the computations made after obtaining the GLIMMIX output.

```
The GLIMMIX Procedure

                        Model Information
Data Set                        WORK.B
Response Variable               y
Response Distribution           Gaussian
Link Function                   Identity
Variance Function               Default
Variance Matrix                 Not blocked
Estimation Technique            Restricted Maximum Likelihood
Degrees of Freedom Method       Containment
  Class Level Information
Class     Levels     Values
trt            2     1 2
rep            4     1 2 3 4
Number of Observations Read          16
Number of Observations Used          16
             Dimensions
G-side Cov. Parameters         1
R-side Cov. Parameters         1
Columns in X                   3
Columns in Z                   8
Subjects (Blocks in V)         1
Max Obs per Subject           16

            Parameter Search

                               Objective
   CovP1        CovP2           Function

  5.0000      10.0000     66.284236398

             Optimization Information

Optimization Technique          Dual Quasi-Newton
Parameters in Optimization      2
Equality Constraints            2
Lower Boundaries                2
Upper Boundaries                2
Fixed Effects                  Profiled
```

```
Starting From                 Data

                                Iteration History
                                        Objective                   Max
Iteration  Restarts  Evaluations        Function        Change Gradient
    0          0          4           66.284236398                0

      Convergence criterion (ABSGCONV=0.00001) satisfied.

           Fit Statistics
-2 Res Log Likelihood          66.28
AIC  (smaller is better)        66.28
AICC (smaller is better)        66.28
BIC  (smaller is better)        66.28
CAIC (smaller is better)        66.28
HQIC (smaller is better)        66.28
Generalized Chi-Square           0.00
Gener. Chi-Square / DF           0.00

 Covariance Parameter Estimates
                        Standard
Cov Parm    Estimate       Error
rep(trt)      5.0000           .
Residual     10.0000           .

        Type III Tests of Fixed Effects
              Num      Den
Effect        DF       DF    F Value    Pr > F
trt             1        6       5.00    0.0667

                  Num
Obs  Effect  DF  DenDF  FValue  ProbF   alpha  ncparm  F_critical power
1    trt     1   6      5.00    0.0667  0.05   5       5.98738    0.46741
```

Results and Conclusions

The predicted power for this future hypothetical design is 0.47. Increasing the number of replicates or samples would increase the predicted power, a result that can easily be investigated by repeated runs of this code, simply by changing the values in the two "do" statements.

EXERCISE #2.
Solution

Expanding on Example #1, we broaden the SAS code to a number of different designs using a SAS macro.

Example

Consider the proposed completely randomized design in which treatments are replicated as shown in Figure 1C and, additionally, multiple sampling (observational) units are created within each experimental unit, from which one data point is collected on each observational unit (Example #1). Replicates are nested within treatments and sampling units are nested within experimental units. The goal remains to detect a difference between treatment means of 5% at a Type 1 error rate of $\alpha = 0.05$. Prior estimates of experimental error (5) and sampling error (10) are available. We wish to predict the power for six different designs with $r = 4$, 5, or 6 replicates and $s = 2$ or 3 sampling units per replicate.

SAS code

The SAS code below computes the expected power for the design described above for all six variations of the numbers of replicates per treatment and sampling units per experimental unit. It first creates a macro titled "one". The macro then creates two new variables: "obsv" is the number of sampling units per experimental unit and "repl" is the number of replicates. These two values are allowed to vary with maxima of "obsmax" and "repmax", respectively. The line that reads "%one(3,6);" is the place to set the upper limits for the investigation. In this example, we have chosen to vary the number of replicates from 4 to 6 and the number of sampling units from 2 to 3. The remainder of the code is identical to that in Example #1.

```
options nocenter;
%macro one(obsmax,repmax);
data a;
%do obsv=2 %to &obsmax;
group1=&obsv;
%do repl=4 %to &repmax;
group2=&repl;
do obs=1 to &obsv by 1;
do rep=1 to &repl by 1;
do trt=0 to 1 by 1;
output; end; end; end;
%end; %end;
%mend one;
%one(3,6);          /*   <--- change values here     */
run;
proc sort; by group1 group2;
data b; set a; by group1 group2;
if trt=0 then y=95;
if trt=1 then y=100;
run;
proc glimmix; class trt rep; by group1 group2;
model y = trt;
random rep(trt);
parms (5)(10) / hold=1,2;
ods output tests3=power_terms;
data power; set power_terms;
alpha=0.05;
ncparm=numdf*Fvalue;
F_critical=finv(1-alpha, numdf, dendf, 0);
power=1-probf(F_critical, numdf, dendf, ncparm);
proc print; run;
```

SAS output

The output below is abbreviated by eliminating all the proc glimmix output, which is necessary only for the purpose of checking the SAS run for errors. The output below is the result of the "proc print" statement at the end of the SAS code, printing all the calculated parameter estimates for the six design scenarios, shown under the headings "group1" (number of sampling units) and "group2" (number of replicates).

Obs	group1	group2	Effect	Num DF	DenDF	FValue	ProbF	alpha	ncparm	F_critical	power
1	2	4	trt	1	6	5.00	0.0667	0.05	5.00	5.98738	0.46741
2	2	5	trt	1	8	6.25	0.0369	0.05	6.25	5.31766	0.59308
3	2	6	trt	1	10	7.50	0.0209	0.05	7.50	4.96460	0.69494
4	3	4	trt	1	6	6.00	0.0498	0.05	6.00	5.98738	0.53734
5	3	5	trt	1	8	7.50	0.0255	0.05	7.50	5.31766	0.67085
6	3	6	trt	1	10	9.00	0.0133	0.05	9.00	4.96460	0.77140

Results and Conclusions

The output allows direct comparisons of designs that are both statistically and logistically efficient. It shows that there are multiple ways to achieve an expected level of power, e.g. $r = 6$ replicates and $s = 2$ sampling units is roughly equivalent to $r = 5$ replicates and $s = 3$ sampling units. The results can be expanded to a wider range of values and used to graphically display the design comparisons as shown in Figure 2, in which $r = 3$ to 20 and $s = 3$ to 20.

EXERCISE #3.
Solution:

The exercise is similar to Exercise #2, but the design differs, providing another illustration for conducting power analyses.

Example

The experiments described in Casler (1998; 2013) were used to obtain the following estimates of random factors: blocks (0), treatment ´ location interaction (0.02), and residual variance (2.0). Power was predicted for a randomized complete block design, but the random block effect was assumed to be zero, based on previous estimates from (Casler, 1998). The desired detection limit was set to 5%, with representative treatment means of 9.5 and 10 with a Type 1 error rate of $\alpha = 0.05$.

SAS code

The SAS code below computes the expected power for the design described above for 25 variations of the numbers of replicates per location and number of locations. It first creates a macro entitled "two". The macro then creates two new variables: "locn" is the number of locations and "repl" is the number of replicates. These two values are allowed to vary with maxima of "locmax" and "repmax", respectively. The line that reads "%two(6,8)" is the place to set the upper limits for the investigation. In this example, we have chosen to vary the number of replicates from 4 to 8 and the number of locations from 2 to 6. The remainder of the code is similar to that in Examples #1 and 2.

```
options nocenter;
%macro two(locmax,repmax);
data a;
%do locn=2 %to &locmax;
group1=&locn;
%do repl=4 %to &repmax;
group2=&repl;
do loc=1 to &locn by 1;
do rep=1 to &repl by 1;
do trt=0 to 1 by 1;
output;
end;
end;
end;
%end;
%end;
%mend two;
%two(6,8);           /*   <--- change here     */
run;
proc sort; by group1 group2;
```

```
data b; set a; by group1 group2;
if trt=0 then y=9.5;
if trt=1 then y=10;
run;
proc glimmix; class loc trt rep; by group1 group2;
model y = trt;
random trt*loc;
parms (0.02)(0.2) / hold=1,2;
ods output tests3=power_terms;
data power;
set power_terms;
alpha=0.05;
ncparm=numdf*Fvalue;
F_critical=finv(1-alpha, numdf, dendf, 0);
power=1-probf(F_critical, numdf, dendf, ncparm);
proc print;
run;
```

SAS output

The output below is abbreviated by eliminating all the proc glimmix output, which is necessary only for the purpose of checking the SAS run for errors. The output below is the result of the "proc print" statement at the end of the SAS code, printing all the calculated parameter estimates for the six design scenarios, shown under the headings "group1" (number of locations) and "group2" (number of replicates).

Obs	group1	group2	Effect	DF	DenDF	FValue	ProbF	alpha	ncparm	F_critical	power
1	2	4	trt	1	2	3.57	0.1994	0.05	3.5714	18.5128	0.20180
2	2	5	trt	1	2	4.17	0.1780	0.05	4.1667	18.5128	0.22463
3	2	6	trt	1	2	4.69	0.1628	0.05	4.6875	18.5128	0.24407
4	2	7	trt	1	2	5.15	0.1514	0.05	5.1471	18.5128	0.26082
5	2	8	trt	1	2	5.56	0.1425	0.05	5.5556	18.5128	0.27539
6	3	4	trt	1	4	5.36	0.0816	0.05	5.3571	7.7086	0.42377
7	3	5	trt	1	4	6.25	0.0668	0.05	6.2500	7.7086	0.47726
8	3	6	trt	1	4	7.03	0.0569	0.05	7.0312	7.7086	0.52112
9	3	7	trt	1	4	7.72	0.0499	0.05	7.7206	7.7086	0.55749
10	3	8	trt	1	4	8.33	0.0447	0.05	8.3333	7.7086	0.58799
11	4	4	trt	1	6	7.14	0.0369	0.05	7.1429	5.9874	0.60896
12	4	5	trt	1	6	8.33	0.0278	0.05	8.3333	5.9874	0.67420
13	4	6	trt	1	6	9.37	0.0222	0.05	9.3750	5.9874	0.72379
14	4	7	trt	1	6	10.29	0.0184	0.05	10.2941	5.9874	0.76214
15	4	8	trt	1	6	11.11	0.0157	0.05	11.1111	5.9874	0.79231
16	5	4	trt	1	8	8.93	0.0174	0.05	8.9286	5.3177	0.74472
17	5	5	trt	1	8	10.42	0.0121	0.05	10.4167	5.3177	0.80635
18	5	6	trt	1	8	11.72	0.0090	0.05	11.7187	5.3177	0.84922
19	5	7	trt	1	8	12.87	0.0071	0.05	12.8676	5.3177	0.87980
20	5	8	trt	1	8	13.89	0.0058	0.05	13.8889	5.3177	0.90214
21	6	4	trt	1	10	10.71	0.0084	0.05	10.7143	4.9646	0.83824
22	6	5	trt	1	10	12.50	0.0054	0.05	12.5000	4.9646	0.88893
23	6	6	trt	1	10	14.06	0.0038	0.05	14.0625	4.9646	0.92097
24	6	7	trt	1	10	15.44	0.0028	0.05	15.4412	4.9646	0.94192
25	6	8	trt	1	10	16.67	0.0022	0.05	16.6667	4.9646	0.95608

Results and Conclusions

The output allows any researcher to make direct comparisons of designs that are both statistically and logistically efficient. The results can be expanded to a wider range of values and used to graphically display the design comparisons as shown in Figure 2 for $l = 2$ to 6 locations and $r = 4$ to 20 replicates.

CHAPTER 5: MULTIPLE COMPARISON PROCEDURES: THE INS AND OUTS

David J. Saville

Solutions

Exercise 1

Each of the main effect means for the "within row spacing" factor is an average of 12 data values, so each has an effective sample size of $n = 12$. The LSD(5%) which is appropriate for comparing the two main effect means is therefore

$$LSD(5\%) = 2.131 \times \sqrt{[2 \times 853{,}113 / 12]} = 804$$

where 2.131 is the two-sided 5% critical value for the t distribution with the residual degrees of freedom (15) (this critical value for t_{15} can also be calculated in Excel by typing the formula "=tinv(0.05,15)" into any cell), and 853,113 is the residual mean square (which is also the "pooled variance estimate").

Exercise 2.

(a) When sorted into descending order, the treatment means are 7604, 7493, 6150, 5679, 4838 and 4192. We now search for "homogeneous" groups of means, and assign a letter to each such group.

We start our search with the largest mean, 7604. The second largest mean, 7493, differs from 7604 by only 111, which is less than the least significant difference (LSD(5%)=1392), so the two means do not differ significantly, so we include this mean of 7493 in a homogeneous group along with the first mean of 7604. The third largest mean, 6150, however, differs from 7604 by 1454, which is greater than the LSD(5%) of 1392, so the two means differ significantly, so we cannot include this mean of 6150 in a homogeneous group along with the first two means. Therefore our first homogeneous group consists of just the first two means, 7604 and 7493. To indicate this result, we assign the letter "a" to each of these means.

We now continue our search for homogeneous groups by forgetting about the largest mean, and examining the second largest mean, 7493, in relation to the remaining four means. Now the third largest mean, 6150, differs from 7493 by 1343, which is less than the LSD(5%) of 1392, so the two means do not differ significantly, so we include this mean of 6150 in a homogeneous group along with the mean of 7493. The fourth largest mean, 5679, however, differs from 7493 by 1814, which is greater than the LSD(5%) of 1392, so the two means differ significantly, so we cannot include this mean of 5679 in a homogeneous group along with the other two means. Therefore our second homogeneous group consists of just two means, 7493 and 6150. To indicate this result, we assign the letter "b" to each of these means.

Continuing our search, we forget about the two largest means, and examine the third largest mean, 6150, in relation to the remaining three means. The fourth largest mean, 5679, differs from 6150 by 471, which is less than the LSD(5%) of 1392, so the two means do not differ significantly, so we include this mean of 5679 in a homogeneous group along with the mean of 6150. The fifth largest mean, 4838, differs from 6150 by 1312, which is also less than the LSD(5%) of 1392, so the two means do not differ significantly, so we also include this mean of 4838 in a homogeneous group along with the other two means. The sixth largest mean, 4192, however, differs from 6150 by 1958, which is greater than the LSD(5%) of 1392, so the two means differ significantly, so we cannot include this mean of 4192 in a homogeneous group along with the other three means. Therefore our third homogeneous group consists of three means, 6150, 5679 and 4838. To indicate this result, we assign the letter "c" to each of these three means.

Continuing our search, we forget about the three largest means, and examine the fourth largest mean, 5679, in relation to the remaining two means. The fifth largest mean, 4838, differs from 5679 by 841, which is less than the LSD(5%) of 1392, so the two means do not differ significantly, so we include the mean of 4838 in a homogeneous group along with the mean of 5679. The sixth largest mean, 4192, however, differs from 5679 by 1487, which is greater than the LSD(5%) of 1392, so the two means differ significantly, so we cannot include the mean of 4192 in a homogeneous group along with the other two means. Therefore our fourth homogeneous group consists of two means, 5679 and 4838. We notice, however, that this fourth homogeneous group is included in the third homogeneous group (assigned the letter "c"), so we do not declare a fourth homogeneous group, and do not assign any more letters to these two means.

To complete our search, we compare the fifth largest mean, 4838, to the only remaining mean, 4192. These means differ by 646, which is less than the LSD(5%) of 1392, so the two means do not differ significantly, so we include the mean of 4192 in a homogeneous group along with the mean of 4838. Therefore our fourth homogeneous group consists of two means, 4838 and 4192. To indicate this result, we assign the letter "d" to each of these three means.

This completes the process. The final result, in terms of the sorted means, is:

7604	a
7493	ab
6150	bc
5679	c
4838	cd
4192	d

When re-sorted into the order of the treatments, this gives the same lettering as shown in Table 2.

Aside: In this example, the six treatments were equally replicated, so a single LSD(5%) could be used for comparing all pairs of treatment means. The above procedure can then be easily performed by computer. However, if the treatments had been unequally replicated, several different LSD(5%) values would have been required and the procedure is more complicated, with the result that attempts at generating an appropriate computer routine are not universally successful.

(b) For the within-row spacing of 5 cm, *yes*, there is a 5% significant difference between the "3 rows per bed" treatment mean (6150) and the "5 rows per bed" treatment mean (7604), since they do not have a letter in common (which reflects the fact that they differ by 1454, which is greater than the LSD(5%) of 1392). Similarly, for the within-row spacing of 10 cm, *yes*, there is a 5% significant difference between the “3 rows per bed” treatment mean (4192) and the “5 rows per bed” treatment mean (5679), since they do not have a letter in common (which reflects the fact that they differ by 1487, which is greater than the LSD(5%) of 1392).

For each within-row spacing (5 cm and 10 cm), *no*, there is no significant difference between the “4 rows per bed” treatment mean and either the “3 rows per bed” or “5 rows per bed” treatment means, since for all four pairwise comparisons the two means being compared have a letter in common (which reflects the fact that the two means always differ by less than the LSD(5%) of 1392).

(c) For each between-row spacing (3, 4, and 5 rows per bed), yes, there is a 5% significant difference between the “5 cm” treatment mean and the “10 cm” treatment mean? For example, the two “3 rows per bed” treatment means are 6150 and 4192 for 5 cm and 10 cm spacing, respectively, and these means differ significantly since they do not have a letter in common (which reflects the fact that they differ by 1958, which is greater than the LSD(5%) of 1392). For “4 rows per bed” the respective means again do not have a letter in common, and differ by 2655 (>1392). For “5 rows per bed” the respective means also do not have a letter in common, and differ by 1925 (>1392).

CHAPTER 6: LINEAR REGRESSION TECHNIQUES

Christel Richter and Hans-Peter Piepho

Appendix 1 (Refers to Example 1: Datafile EEL with variables LENGTH and WEIGHT)

Grey: Externally Studentized residuals with $|\hat{e}_i^{**}| > 2$. Framed: high leverage

Eel_no	Length (cm)	Weight (g)	Eel_no	Length (cm)	Weight (g)
1	33	108.6	18	46	184.8
2	34	114.1	19	46	191.8
3	36	120.4	20	46	189.7
4	36	128.6	21	47	202.6
5	37	137.5	22	47	198.9
6	39	144.2	23	47	198.2
7	39	148.3	24	48	209.6
8	40	152.4	25	49	212.1
9	41	160.5	26	51	224.5
10	42	166.4	27	51	224.7
11	42	165.9	28	51	228.3
12	42	162.8	29	51	221.6
13	43	179.0	30	52	231.7
14	43	172.1	31	53	246.2
15	44	178.0	32	54	247.5
16	45	189.7	33	55	254.8
17	46	194.9	34	58	275.0

```
DATA eel;
SET eel;
LABEL length='Length [cm]' weight='Weight [g]';
RUN;

ODS GRAPHICS ON;

TITLE 'Example 1: Generation of a template for Figure 1';
PROC TEMPLATE;
DEFINE STATGRAPH eel_temp;
BEGINGRAPH;
ENTRYTITLE "Weight and length of eels" /TEXTATTRS=(SIZE=11pt);
LAYOUT lattice/COLUMNS = 2 ROWS = 2 COLUMNWEIGHTS = (.8 .2) ROWWEIGHTS
= (.8 .2)
COLUMNDATARANGE = union ROWDATARANGE = union;
COLUMNAXES;
COLUMNAXIS /LABEL = "Length [cm]" GRIDDISPLAY = on
LABELATTRS=(SIZE=12)TICKVALUEATTRS=(SIZE=10);
COLUMNAXIS /LABEL = "" GRIDDISPLAY = on;
ENDCOLUMNAXES;
ROWAXES;
ROWAXIS /LABEL = "Weight [g]" GRIDDISPLAY = on LABELATTRS=(SIZE=12)
TICKVALUEATTRS=(SIZE=10);
ROWAXIS /LABEL = "" GRIDDISPLAY = on;
ENDROWAXES;
LAYOUT overlay;
SCATTERplot X = length Y = weight/MARKERATTRS=(COLOR=black SIZE=10
```

```
SYMBOL=circlefilled);
ENDLAYOUT;
BOXPLOT Y=weight/ORIENT = vertical;
BOXPLOT Y=length/ORIENT = horizontal;
ENDLAYOUT;
ENDGRAPH;
END;
RUN;
TITLE 'Example 1: Figure 1';
PROC SGRENDER DATA = eel TEMPLATE = eel_temp;
RUN; QUIT;

TITLE 'Example 1: Regression with Figure 3 and some further
representations with PROC REG';
PROC REG DATA=eel PLOTS(LABEL)=all;
MODEL weight = length /CLB;
OUTPUT OUT=eel_out P=yhat LCLM=lclm UCLM=uclm LCL=lcl UCL=ucl R=yresid
STUDENT=student RSTUDENT=rstudent H=h COOKD=cookd COVRATIO=covratio
DFFITS=dffits PRESS=PRESS;
RUN; QUIT;

TITLE 'Example 1: Regression with PROC MIXED';
PROC MIXED DATA=eel plots=all;
MODEL weight = length /s CL outp=eel_out_mixed;
RUN; QUIT;

TITLE 'Example 1: Regression with PROC GLM';
PROC GLM DATA=eel plots=all;
MODEL weight = length /SOLUTION CLPARM ;
OUTPUT OUT=eel_out_glm P=yhat LCLM=lclm UCLM=uclm LCL=lcl UCL=ucl
R=yresid STUDENT=student RSTUDENT=rstudent H=h COOKD=cookd
COVRATIO=covratio DFFITS=dffits PRESS=PRESS;
RUN; QUIT;

TITLE 'Example 1: Figure 4 Confidence ellipse';
PROC CORR DATA=eel PLOTS=SCATTER(ellipse=confidence alpha= 0.05);
VAR length weight;
RUN; QUIT;

TITLE 'Example 1: Figure 4 Prediction ellipse';
PROC CORR DATA=eel PLOTS=SCATTER(alpha= 0.05);
VAR length weight;
RUN; QUIT;

TITLE 'Example 1: Figure 6 and tests for normality';
PROC UNIVARIATE DATA= eel_out NORMAL;
QQPLOT student/NORMAL (MU=0 SIGMA=1) ODSTITLE='Q-Q Plot for Eel_
Weight';
VAR student;
RUN; QUIT;

TITLE 'Example 1: Figure 7';
DATA eel_out;
SET eel_out; IF student<-2 OR student>2 THEN studcrit="S obs.
no.="||trim(left(_N_));
IF studcrit ne '      ' THEN crit =length;
RUN; QUIT;

PROC SGPLOT DATA= eel_out NOAUTOLEGEND;
TITLE 'Internally studentized residual against length';
SCATTER X=length Y=student;REFLINE 0 2 -2;
SCATTER X=crit Y=student/ DATALABEL=studcrit
DATALABELATTRS=(Family=Arial SIZE=10
STYLE=Italic Weight=Bold) MARKERATTRS=(COLOR=black SIZE=8
SYMBOL=circlefilled);
REFLINE 0 2 -2;
RUN; QUIT;
```

Appendix 2 (Refers to Example 2: Datafile CALIBRATION with variables *C* and EXT)

Grey: Externally Studentized residuals with $\left|\hat{e}_i^{**}\right| > 2$. No high leverage.

Obs_no	Concentration (mmol/l)	Extinction
1	0.5	55.5
2	1.0	77.0
3	1.5	100.7
4	3.0	165.5
5	5.0	253.4

Regression analysis and Fig. 3 analogous to PROC REG of Example 1; for Fig. 2, the Template of Example 1 must be adjusted (among others: boxplot only for the extinction)

```
TITLE 'Example 2: Generation of concentration values as initial
values for the iteration';
DATA cali_start;
DO start=4.3 TO 4.6 BY 0.01;
OUTPUT; END;
RUN; QUIT;

TITLE 'Example 2: Iterative solution for CL and CU';
/*rootMSE and regression function from PROC REG; Cmean and SSC
from PROC UNIVARIATE*/
PROC MODEL DATA=cali_start OUT=result;
rootMSE=0.7320449; n=5; Cmean=2.2; SSC=13.3; ext=230;
EXT1=33.70602+43.9609*CL; /*regression function*/
EXT2=33.70602+43.9609*CU; /*regression function*/
eq.CL=EXT2- rootMSE *sqrt(1+1/n+(CL- Cmean)**2/
SSC)*tinv(0.975,n-2)-ext;
eq.CU=EXT1+ rootMSE *sqrt(1+1/n+(CU- Cmean)**2/
SSC)*tinv(0.975,n-2)-ext;
SOLVE CL CU;
RUN; QUIT;
```

Appendix 3 (Refers to Example 3 and Example 3 (modfied): Datafile FIBER with variables DAY, CONTENT, and BLOCK)

Grey: Externally Studentized residuals with $\left|\hat{e}_i^{**}\right| > 2$. No high leverage

No. obs.	Day	Crude fiber content (g/kg)	Block	No. obs.	Day	Crude fiber content (g/kg)	Block
1	0	218	1	11	10	289	3
2	0	225	2	12	10	297	4
3	0	229	3	13	15	297	1
4	0	239	4	14	15	307	2
5	5	246	1	15	15	317	3
6	5	258	2	16	15	343	4
7	5	254	3	17	20	316	1
8	5	269	4	18	20	336	2
9	10	257	1	19	20	354	3
10	10	275	2	20	20	351	4

For Fig. 2, the Template of Example 1 must be adjusted; Fig. 3, 6, and 7, Table 2 A and D analogous to the code of Example 1

```
DATA fiber;
SET fiber;
LABEL content='Fiber Content [g/kg]' Day='Days after the first
cut';
RUN; QUIT;

TITLE 'Example 3: Table 2 B';
PROC REG DATA=fiber;
MODEL content=day / lackfit;
RUN; QUIT;

TITLE 'Example 3: Remark to explain the lack-of-fit';
DATA fiber;
SET fiber;
day2=day*day; day3=day2*day; day4=day2*day2;
RUN; QUIT;
PROC REG DATA=fiber;
MODEL content=day day2 day3 day4/lackfit ss1;
RUN; QUIT;

TITLE 'Example 3: Table 2 C with PROC GLM';
PROC GLM DATA=fiber;
CLASS day;
MODEL content=day;
RUN; QUIT;

TITLE 'Example 3: Table 2 C with PROC MIXED';
PROC MIXED DATA=fiber;
CLASS day;
MODEL content=day;
RUN; QUIT;

TITLE 'Example 3: Calculation of the means per day';
PROC MEANS DATA=fiber;
CLASS day;
VAR content;
OUTPUT OUT=fiber_mean_file MEAN=content_mean;
RUN; QUIT;

TITLE 'Example 3: Table 3';
PROC REG DATA=fiber_mean_file;
MODEL content_mean = day;
OUTPUT OUT=fiber_mean_out P=yhat STDP=stdp LCLM=lclmean UCLM=uclmean
STDR=stdr STDI=stdi LCL=lclind UCL=uclind;
RUN; QUIT;

TITLE 'Example 3 (modified): Table 18 A';
PROC GLM data=fiber;
CLASS block;
MODEL content =day block;
RUN; QUIT;

TITLE 'Example 3 (modified): Table 18 B';
PROC MIXED DATA=fiber;
/*Estimated function values and confidence intervals per fixed block are
given in fiber_fix*/
CLASS block;
MODEL content = day block /S CL OUTP=fiber_fix;
ESTIMATE 'b0+block1' int 1 block 1 /cl;
ESTIMATE 'b0+block2' int 1 block 0 1/cl;
ESTIMATE 'b0+block3' int 1 block 0 0 1/cl;
ESTIMATE 'b0+block4' int 1 block 0 0 0 1/cl;
RUN; QUIT;

TITLE 'Example 3 (modified): Table 19 and Table 20 (broad inference';
PROC MIXED DATA=fiber;
```

```
/*Confidence intervals (broad inference) are given in fiber_rand*/
CLASS block;
MODEL content =day /S CL ddfm=Kenwardroger(firstorder) OUTPM=fiber_rand;
RANDOM block /S;
RUN; QUIT;

TITLE 'Example 3 (modified): Table 20';
/*To have a common divisor all values in the estimate statement are
multiplied with 4 = number of blocks*/
PROC MIXED DATA=fiber;
/*Calculation of confidence intervals for the mean of all fixed blocks*/
CLASS block;
MODEL content =day block /S CL;
ESTIMATE 'day0'   int 4 block 1 1 1 1 day  0/divisor=4 cl;
ESTIMATE 'day5'   int 4 block 1 1 1 1 day 20/divisor=4 cl;
ESTIMATE 'day10'  int 4 block 1 1 1 1 day 40/divisor=4 cl;
ESTIMATE 'day15'  int 4 block 1 1 1 1 day 60/divisor=4 cl;
ESTIMATE 'day20'  int 4 block 1 1 1 1 day 80/divisor=4 cl;
RUN; QUIT;

TITLE 'Example 3 (modified): Table 20';
PROC MIXED DATA=fiber;
/*Calculation of confidence intervals for the mean of all random
blocks*/
CLASS block;
MODEL content =day /S CL ddfm=Kenwardroger(firstorder);
RANDOM block /S;
ESTIMATE 'day0'  int 4 day 0  | block 1 1 1 1 /divisor=4 cl;
ESTIMATE 'day5'  int 4 day 20 | block 1 1 1 1 /divisor=4 cl;
ESTIMATE 'day10' int 4 day 40 | block 1 1 1 1 /divisor=4 cl;
ESTIMATE 'day15' int 4 day 60 | block 1 1 1 1 /divisor=4 cl;
ESTIMATE 'day20' int 4 day 80 | block 1 1 1 1 /divisor=4 cl;
RUN; QUIT;
```

Appendix 4 (Refers to Example 4: Datafile GRASS with variables WGRASS and YIELD)

Grey: Externally Studentized residuals with $\left|\hat{e}_i^{**}\right| > 2$. Framed: high leverage

No. Obs.	Rx_i	Wind grass (number/m²) x_i	yield (g / plot) y_i	Ry_i	No. Obs.	Rx_i	Wind grass (number/m²) x_i	yield (g / plot) y_i	Ry_i
1	1.5	0	9310	48	27	27.5	99	6410	23
2	1.5	0	8460	43	28	27.5	99	6640	25.5
3	4	1	9770	52	29	29	100	6010	19
4	4	1	9320	49	30	30	101	6940	28
5	4	1	8620	44	31	31.5	102	4930	14
6	6	2	7850	36.5	32	31.5	102	7620	33
7	7.5	3	9520	51	33	33.5	132	5680	17
8	7.5	3	9080	46	34	33.5	132	7570	32
9	9	4	6340	22	35	35	145	5240	16
10	10	5	9340	50	36	36	152	6320	21
11	11	17	7940	39	37	37	161	3970	9
12	12	21	8730	45	38	38	167	4790	13
13	13	22	7870	38	39	39	197	5230	15
14	14	24	8160	41	40	40	243	5980	18
15	15	31	7700	35	41	41	250	7240	29
16	16	37	8120	40	42	42	251	2340	5
17	17	46	7850	36.5	43	43	258	4320	12
18	18	51	7530	31	44	44	268	2830	6
19	19.5	56	7630	34	45	45	288	3810	8
20	19.5	56	9120	47	46	46	305	4260	11

21	21.5	57	6640	25.5	47	47	311	4130	10
22	21.5	57	7490	30	48	48	337	3050	7
23	23	61	6280	20	49	49	901	1740	2
24	24	81	8450	42	50	50	927	1750	3
25	25	84	6920	27	51	51	1102	1980	4
26	26	88	6550	24	52	52	1204	540	1

For Fig. 2, the Template of Example 1 must be adjusted; for Fig. 3, 6, and 7 the code of Example 1 can be used

```
DATA grass;
SET grass; sqrt_wgrass=sqrt(wgrass); log_yield=log(yield);
LABEL wgrass='wind grass [number/plot]' yield='Yield [g/plot]'
sqrt_wgrass='sqrt(wind grass)';
RUN; QUIT;

TITLE 'Example 4: Spearman's correlation coefficient';
PROC CORR DATA=grass SPEARMAN;
VAR wgrass yield;
RUN; QUIT;

TITLE 'Example 4: Regression with logarithm of yield
(multiplicative errors after back-transformation)';
PROC REG DATA=grass;
MODEL log_yield=wgrass;
RUN; QUIT;

TITLE 'Example 4: Non-linear regression yield=a*exp(b*wgrass)
with additive errors';
PROC NLIN DATA=grass PLOTS=all;
PARMS  a=8000 b=-0.01;
MODEL yield=a*exp(b*wgrass);
RUN; QUIT;

TITLE 'Example 4: Non-linear regression
yield=a*exp(b*wgrass)+c with additive errors and Fig. 8';
PROC NLIN DATA=grass PLOTS=all;
PARMS a=7000 b=-0.01 c=1200;
MODEL yield=a*exp(b*wgrass)+c;
RUN; QUIT;
```

Appendix 5 (Refers to Example 5: Datafile SHAPE with input variables x1, x2, BLOCK, and YIELD)

```
TITLE 'Example 5: DATA input and calculation of the variables
TREATMENT, AREA, SHAPE_INDEX1, and SHAPE_INDEX2';
DATA space;
INPUT x1 x2 @@; treatment=x1*1000+x2; area=x1*x2;
shape_ind1=x1/x2; shape_ind2=(x1+x2)/2/sqrt(x1*x2);
DO block=1 TO 4; INPUT yield@@; OUTPUT; END;
DATALINES;
30 30 5.95 5.30 6.50 6.35
30 24 7.10 6.45 6.60 5.75
30 20 7.00 6.50 6.35 8.90
30 15 8.10 5.50 6.60 7.50
24 24 8.85 7.65 7.00 7.90
24 20 7.65 6.90 8.25 8.30
24 15 7.80 6.75 8.20 7.25
20 20 8.05 6.65 8.10 8.05
20 15 9.30 8.75 8.75 8.00
15 15 9.35 8.10 7.60 7.75
;TITLE 'Example 5: Table 6 A TO D';
PROC GLM DATA=space;
CLASS block treatment;
MODEL yield =block treatment;
```

```
RUN; QUIT;

PROC GLM DATA=space;
CLASS block treatment;
MODEL yield =block area treatment / SS1;
RUN; QUIT;

PROC GLM DATA=space;
CLASS block treatment;
MODEL yield =block area shape_ind1 treatment / SS1;
RUN; QUIT;
PROC GLM DATA=space;
CLASS block treatment;
MODEL yield =block area shape_ind2 treatment / SS1;
RUN; QUIT;

TITLE 'Example 5: Table 7';
PROC GLM DATA=space;
CLASS block;
MODEL yield =block area / SS1;
OUTPUT OUT=residual_B r=resi_B;
RUN; QUIT;

PROC GLM DATA=space;
CLASS block;
MODEL yield =block area shape_ind1 / SS1;
OUTPUT OUT=residual_C r=resi_C;
RUN; QUIT;

PROC GLM DATA=space;
CLASS block;
MODEL yield =block area shape_ind2 / SS1;
OUTPUT OUT=residual_D r=resi_D;
RUN; QUIT;

PROC SORT DATA= residual_B; BY treatment block; RUN;
PROC SORT DATA= residual_C; BY treatment block; RUN;
PROC SORT DATA= residual_D; BY treatment block; RUN;

DATA resi;
MERGE residual_B residual_C residual_D; BY treatment block;
KEEP treatment block yield resi_B resi_C resi_D;
RUN; QUIT;

PROC MEANS DATA=resi;
VAR yield resi_B resi_C resi_D;
BY treatment;
OUTPUT OUT=mwresi MEAN= mwyield mwresi_B mwresi_C mwresi_D;
RUN; QUIT;
```

Appendix 6 (Refers to Example 6: Datafile POTATO with variables SIZE and WEIGHT).

The datafile Potato.xls is available in the supplemental material.

```
DATA potato;
SET potato;
size2=size*size; size3=size*size*size;
size_reciprocal = 1/size;
label size='Size [mm]' weight='Weight [g]';
RUN; QUIT;

TITLE 'Example 6: Table 8 A and B';
/* Sequential approach with sequence x1→ x2 → x3; partially also
with PROC REG (with option ss1) and PROC MIXED (with option htype=1)
possible*/
```

```
PROC GLM DATA= potato;
MODEL weight = size size2 size3 /SOLUTION SS1;
RUN; QUIT;
/* Sequential approach with sequence x3→ x1 → x2*/
PROC GLM DATA= potato;
MODEL weight =size3 size size2 / SOLUTION SS1;
RUN; QUIT;
/* Partial approach does not depend on the sequence*/
PROC GLM DATA= potato;
MODEL weight =size size2 size3 / SOLUTION ;
RUN; QUIT;

TITLE 'Example 6: Table 8 C';
/*pcorr2 bases on the partial approach, pcorr1 on the sequential
approach*/
PROC REG DATA=potato;
MODEL weight =size size2 size3/ PCORR2 TOL;
RUN; QUIT;

TITLE 'Example 6: Table 9 A';
/*m123... m3 are labels for the different models with intercepts. The
fit criteria are in the potato_info_int file*/
PROC REG DATA= potato OUTEST= potato_info_int;
m123: MODEL weight = size size2 size3 / ADJRSQ AIC PRESS BIC SBC SSE;
m12:  MODEL weight = size size2 / ADJRSQ AIC PRESS BIC SBC SSE;
m13:  MODEL weight = size size3 / ADJRSQ AIC PRESS BIC SBC SSE;
m23:  MODEL weight = size2 size3 / ADJRSQ AIC PRESS BIC SBC SSE;
m1:   MODEL weight = size / ADJRSQ AIC PRESS BIC SBC SSE;
m2:   MODEL weight = size2 / ADJRSQ AIC PRESS BIC SBC SSE;
m3:   MODEL weight = size3 / ADJRSQ AIC PRESS BIC SBC SSE;
RUN; QUIT;

TITLE 'Example 6: Table 9 B';
/*m123... m3 are labels for the different models without
intercepts. The fit criteria are in the potato_info_noint file*/
PROC REG DATA= potato OUTEST= potato_info_noint;
m123: MODEL weight = size size2 size3 / NOINT ADJRSQ AIC PRESS
BIC SBC SSE;
m12:  MODEL weight = size size2 / NOINT ADJRSQ AIC PRESS BIC
SBC SSE;
m13:  MODEL weight = size size3 / NOINT ADJRSQ AIC PRESS BIC
SBC SSE;
m23:  MODEL weight = size2 size3 / NOINT ADJRSQ AIC PRESS BIC
SBC SSE;
m1:   MODEL weight = size / NOINT ADJRSQ AIC PRESS BIC SBC SSE;
m2:   MODEL weight = size2 / NOINT ADJRSQ AIC PRESS BIC SBC SSE;
m3:   MODEL weight = size3 / NOINT ADJRSQ AIC PRESS BIC SBC SSE;
RUN; QUIT;

TITLE 'Example 6: Table 12';
/*first step: PROC TRANSREG to get the LLtransreg-values*/
/*with log(size) as regressor and using lambda=0 */
PROC TRANSREG details DATA= potato PLOTS=all SS2 PBOXCOXTABLE
CL;
MODEL BOXCOX(weight /  LAMBDA=0 ) = log(size);
RUN; QUIT;

/*with log(size) as regressor and searching for the optimal
lambda */
PROC TRANSREG details DATA= potato PLOTS=all SS2 PBOXCOXTABLE
CL;
MODEL BOXCOX(weight /  LAMBDA=0 to 1 by 0.01 ) = log(size);
RUN; QUIT;

/*with size as regressor and using lambda=1/3 if it is in
confidence interval of the optimal lambda*/
PROC TRANSREG details DATA= potato PLOTS=all SS2 PBOXCOXTABLE CL;
```

```
MODEL BOXCOX (weight / CONVENIENT CLL =0.333333 lambda=0.3 TO 0.4 by
0.001) = IDENTITY(size);
RUN; QUIT;

/*second step: Calculation of -2LL=LLnew in Table 12 using the
LLtransreg-values of PROC TRANSREG; for the first row in Table
12: -1166.3, for the second row: -1157.06, and for the third row:
-1159.38*/
DATA calc;
LLtransreg=-1166.3;
n=524; PI=constant('PI'); LLnew=-2*(LLtransreg-n/2*log(2*PI)-(n-2)/2);
PROC PRINT DATA=calc;
RUN; QUIT;

/*for the fourth row in Table 12; with size3 as regressor, no
transformation*/
PROC MIXED data=potato method=ML;
MODEL weight=size3;
run;quit;

TITLE 'Example 6: Table 14';
/*For using the ML-method, replace method = REML (default) with method
= ML */
/*Weighted regression with 1/size*/
PROC MIXED DATA=potato METHOD=REML PLOTS=all;
MODEL weight = size3 /NOINT S CL DDFM=KenwardRoger(firstorder);
WEIGHT size_reciprocal;
RUN; QUIT;

/*Power-of-x model; due to convergence problems, initial values for
the covariance parameters are specified*/
PROC MIXED DATA=potato METHOD=REML PLOTS=all;
MODEL weight = size3 /NOINT S CL DDFM=KenwardRoger(firstorder);
REPEATED / LOCAL=exp(size);
PARMS (0.2) (1.5);
RUN; QUIT;

/*Power-of-mean model in two steps*/
ODS OUTPUT SolutionF=sf;
PROC MIXED DATA=potato;
MODEL weight = size3 /NOINT s;
RUN; QUIT;

PROC MIXED DATA=potato method=REML PLOTS=all;
MODEL weight = size3 /NOINT s;
REPEATED / LOCAL=pom(sf) DDFM=KenwardRoger(firstorder);
RUN; QUIT;

/*Model with individual variance per size; with Fig. 13 right*/
PROC MIXED DATA=potato METHOD=REML PLOTS=all;
MODEL weight = size3 / NOINT s DDFM=KenwardRoger(firstorder)
OUTPM=potato_out_ind residual;
REPEATED / GROUP=size;
RUN; QUIT;

/*Fig. 13*/
PROC REG DATA= potato;
MODEL weight = size3 /NOINT;
OUTPUT OUT=potato_out P=yhat R=yresid STUDENT=student
RSTUDENT=rstudent;
RUN; QUIT;

PROC SGPLOT DATA=potato_out;
TITLE 'Unweighted Regression';
LABEL size='size';
SCATTER x=size y=student;
RUN; QUIT;
```

```
PROC SGPLOT DATA= potato_out_ind;
TITLE 'Regression with individual variance per size';
LABEL size='size' ;
SCATTER x=size y=StudentResid;
RUN; QUIT;
```

Appendix 7 (Refers to Example 7: Datafile APPLE with variables VARIETY, YEAR1_4, and Year1_10)

Grey: Externally Studentized residuals with $\left|\hat{e}_i^{**}\right| > 2$. Framed: high leverage

Variety A Yield (kg/tree)				Variety B Yield (kg/tree)			
Year1_4	Year1_10	Year1_4	Year1_10	Year1_4	Year1_10	Year1_4	Year1_10
22	109	39	137	23	73	40	93
27	119	40	142	27	79	40	96
28	125	41	148	27	72	40	91
30	115	42	154	29	79	41	99
33	127	42	152	30	76	42	87
34	133	42	160	30	86	42	92
34	148	42	155	31	78	44	91
34	141	43	144	32	84	45	97
36	141	43	161	33	79	48	99
36	134	44	170	35	85	51	100
37	144	45	165	35	86	53	105
37	140	47	165	36	82	54	101
38	133	48	164	37	91	55	110
38	144	48	167	37	87	56	108
39	140	54	190	38	89	56	113

```
DATA apple;
SET apple;
LABEL year1_10='Yield [kg/tree] year 1 to 10' year1_4='Yield
[kg/tree] year 1 to 4';
RUN; QUIT;

TITLE 'Example 7: Table 15';
/*Before running the following procedures, the datafile must be
sorted by variety*/
PROC REG DATA=apple;
MODEL year1_10=year1_4 / CLB ADJRSQ;
BY variety;
OUTPUT OUT=apple_out P=yhat LCLM=lclmean UCLM=uclmean
STDR=stdr STDI=stdi  LCL=lclind UCL=uclind;
RUN; QUIT;

TITLE 'Example 7: Fig. 14';
PROC SGPLOT DATA=apple;
TITLE 'Cumulative yield after 10 years against after 4 years';
REG x=year1_4 y=year1_10 /group=variety clm cli
MARKERATTRS=(SIZE=9 SYMBOL=circle);
XAXIS LABEL ="yield [kg/tree] year 1 to 4 "
LABELATTRS=(SIZE=12)  VALUEATTRS=(SIZE=10);
YAXIS LABELATTRS=(SIZE=12) VALUEATTRS=(SIZE=10) VALUES=(50
100 150  200);
RUN; QUIT;

TITLE 'Example 7: Table 16 A';
PROC GLM DATA=apple; /*with PROC GLM*/
```

```
MODEL year1_10=year1_4 /SOLUTION CLPARM;
RUN; QUIT;
PROC MIXED DATA=apple; /*alternatively with PROC MIXED*/
MODEL year1_10=year1_4 /S CL;
RUN; QUIT;

TITLE 'Example 7: Table 16 B';
PROC GLM DATA=apple;/*with PROC GLM*/
CLASS variety;
MODEL year1_10=year1_4 variety variety*year1_4 /SOLUTION CLPARM;
ESTIMATE 'int+b01' intercept 1 variety 1;
ESTIMATE 'int+b02' intercept 1 variety 0 1;
ESTIMATE 'b1+b11' year1_4 1 year1_4*variety 1;
ESTIMATE 'b1+b12' year1_4 1 year1_4*variety 0 1;
ESTIMATE 'b01-b02' variety 1 -1;
ESTIMATE 'b11-b12' year1_4*variety 1 -1;
RUN; QUIT;
PROC MIXED DATA=apple; /*alternatively with PROC MIXED*/
CLASS variety;
MODEL year1_10=year1_4 variety year1_4*variety /S CL;
ESTIMATE 'int+b01' int 1 variety 1 / CL;
ESTIMATE 'int+b02' int 1 variety 0 1 / CL;
ESTIMATE 'b1+b11' year1_4 1 year1_4*variety 1 / CL;
ESTIMATE 'b1+b12' year1_4 1 year1_4*variety 0 1 / CL;
ESTIMATE 'b01-b02' variety 1 -1 / CL;
ESTIMATE 'b11-b12' year1_4*variety 1 -1 / CL;
RUN; QUIT;

TITLE 'Example 7: Table 16 C';
PROC MIXED DATA=apple PLOT=all ;
CLASS variety;
MODEL year1_10= year1_4 year1_4*variety/S CL RESIDUAL OUTP=apple_
OUT_c_varhom;
ESTIMATE 'b1+b11' year1_4 1 year1_4*variety 1;
ESTIMATE 'b1+b12' year1_4 1 year1_4*variety 0 1;
RUN; QUIT;

TITLE 'Example 7: Table 17 A and B';
/*Variance Homogeneity*/
PROC MIXED DATA=apple plot=all ;
CLASS variety;
MODEL year1_10= year1_4 year1_4*variety/S CL RESIDUAL OUTP=apple_
out_c_varhom;
RUN; QUIT;
/*Variance Heterogeneity*/
PROC MIXED DATA=apple plot=all ;
CLASS variety;
MODEL year1_10= year1_4 year1_4*variety/S CL RESIDUAL OUTP=apple_
out_c_varhet;
REPEATED / GROUP=variety;
RUN; QUIT;

TITLE 'Example 7: Figure 15 upper part';
PROC SGPLOT DATA=apple NOAUTOLEGEND;
TITLE 'Cumulative yield after 10 years against after 4 years';
REG x=year1_4 y=year1_10 /clm cli;
SCATTER x=year1_4 y=year1_10/ DATALABEL=variety
DATALABELATTRS=(Family=Arial SIZE=10 STYLE=Italic WEIGHT=Bold)
MARKERATTRS=(SIZE=7 SYMBOL=circlefilled);
XAXIS LABEL ="yield [kg/tree] year 1 to 4 " LABELATTRS=(SIZE=12)
VALUEATTRS=(SIZE=10);
YAXIS LABELATTRS=(SIZE=12) VALUEATTRS=(SIZE=10) VALUES=(50 100 150
200);
RUN; QUIT;

TITLE 'Example 7: Figure 15 lower part left';
PROC SGPLOT DATA=Apple_out_c_varhom;
```

```
TITLE 'Cumulative yield after 10 years against after 4 years';
SCATTER x = year1_4 y = year1_10 /GROUP=variety MARKERATTRS=(SIZE=9
symbol=circle);
BAND x=year1_4 LOWER=LOWER UPPER=UPPER/GROUP=variety
FILLATTRS=(TRANSPARENCY=0);
SERIES x=year1_4 y=pred/GROUP=variety LINEATTRS=(PATTERN=solid
THICKNESS=2);
XAXIS LABEL ="yield [kg/tree] year 1 to 4 " LABELATTRS=(SIZE=12)
VALUEATTRS=(SIZE=10);
YAXIS LABELATTRS=(SIZE=12) VALUEATTRS=(SIZE=10) VALUES=(50  100  150
200);
RUN;QUIT;

TITLE 'Example 7: Figure 15 lower part right';
PROC SGPLOT DATA= apple_OUT_c_varhom NOAUTOLEGEND;
TITLE 'Externally Studentized residuals';
SCATTER x=pred y=studentresid/ DATALABEL=variety
DATALABELATTRS=(Family=Arial SIZE=10 STYLE=Italic WEIGHT=Bold)
MARKERATTRS=(SIZE=7 SYMBOL=circlefilled) ; REFLINE 0;
XAXIS LABEL ="PREDICTED value year 1 TO 10 " LABELATTRS=(SIZE=12)
VALUEATTRS=(SIZE=10);
YAXIS LABELATTRS=(SIZE=12) VALUEATTRS=(SIZE=10) ;
RUN; QUIT;
```

Example 7, Figure 16 left and right analogous to Example 7, Figure 15 lower part with Datafile = apple_out_c_varhet

Appendix 8 (Refers to Example 8: Datafile AIRTEMP with variables TEMP and YEAR)

Grey: Externally Studentized residuals with $\left|\hat{e}_i^{**}\right| > 2$. No high leverage.

Year	Mean air temperature (°C)
1960	8.94
1961	9.35
1962	7.98
1963	7.79
1964	8.42
1965	8.02
1966	8.93
1967	9.50
1968	8.65
1969	7.78
1970	7.93
1971	8.89
1972	8.16
1973	8.53
1974	9.46
1975	9.47
1976	8.51
1977	9.05
1978	8.20
1979	8.00
1980	7.56
1981	8.42
1982	9.42
1983	9.48
1984	8.58
1985	8.00
1986	8.27

Year	Mean air temperature (°C)
1987	7.57
1988	9.57
1989	9.94
1990	10.11
1991	8.85
1992	9.66
1993	8.79
1994	9.85
1995	9.17
1996	7.34
1997	9.04
1998	9.46
1999	10.00
2000	10.41
2001	9.32
2002	9.82
2003	9.46
2004	9.04
2005	9.31
2006	9.89
2007	10.42
2008	10.06
2009	9.43
2010	8.02
2011	10.14
2012	9.27
2013	9.34

```
TITLE 'Example 8: Table 21, Figure 17 and Durbin-Watson test';
PROC REG DATA=airtemp PLOTS (LABEL)=all;
ID Year;
MODEL Temp=Year /DWPROB CLB CLI CLM ADJRSQ;
RUN; QUIT;

TITLE 'Example 8: Table 22';
PROC MIXED DATA=airtemp PLOTS=residualpanel(unpack) ;
MODEL Temp=Year /S CL DDFM=kr(firstorder);
REPEATED / TYPE=AR(1) SUBJECT=intercept;
RUN; QUIT;

TITLE 'Example 8: Analysis as AR(1)';
PROC AUTOREG DATA=airtemp /*AR(1) */;
MODEL Temp=Year /NLAG=1 DW=1 DWPROB;
OUTPUT OUT=auto_ar1 P=pred PM=predm R=rest LCL=lcl LCLM=lclm UCL=ucl
UCLM=uclm;
RUN; QUIT;

TITLE 'Example 8: Analysis as autoregressive model with backward
algorithm';
PROC AUTOREG DATA=airtemp /*Backward algorithm results in AR(4) */;
MODEL Temp=Year /NLAG=12 DW=12 BACKSTEP DWPROB;
OUTPUT OUT=auto_ar4 P=pred PM=predm R=rest LCL=lcl LCLM=lclm UCL=ucl
UCLM=uclm;
RUN; QUIT;

TITLE 'Example 8: Figure 18 at top left';
PROC SGPLOT DATA=auto_ar1;
TITLE 'Analysis by PROC AUTOREG';
BAND X=year LOWER=lcl UPPER=ucl/ TRANSPARENCY=0.6
LEGENDLABEL="Prediction interval" NAME="band1";
BAND X=year LOWER=lclm UPPER=uclm/ TRANSPARENCY =0.3
LEGENDLABEL="Confidence interval" NAME="band2";
SCATTER x=year y=temp/ MARKERATTRS=(SIZE=7 SYMBOL=circlefilled) ;
SERIES x=year y=predm / LINEATTRS=(COLOR=verydarkblue THICKNESS=2 ) ;
SERIES x=year y=pred / LINEATTRS=(COLOR=verydarkred THICKNESS =2 ) ;
XAXIS LABEL ="Year " LABELATTRS=(SIZE=12)  valueATTRS=(SIZE=10);
YAXIS LABEL ="Temperature [grd C]" LABELATTRS=(SIZE=12)
VALUEATTRS=(SIZE=10) ;
KEYLEGEND "band1" "band2" / LOCATION=inside
POSITION=bottomright;
RUN; QUIT;

/*Example 8: Figure 18 at top right analogous to Figure 18 at top left
with datafile = auto_ar4*/
TITLE 'Example 8: Figure 18 at bottom';
PROC NLIN DATA=airtemp PLOTS=all;
PARMS a=7 b=1 c=1960 d=11 e=0.03 ;
MODEL temp=a-b*sin((year-c)*2*3.14/d)+e*year;
OUTPUT OUT=NLIN PREDICTED=pred L95=lcl L95M=lclm U95=ucl U95M=uclm;
RUN; QUIT;

TITLE 'Example 8: Figure 19 ';
PROC NLIN DATA=airtemp PLOTS=all CONVERGEPARM=1E-7;
PARMS a=0.034 b=-60 Year0=1985 ;
IF (Year >Year0) THEN MODEL Temp=a*Year + b;
ELSE MODEL Temp=a*Year0+b;
RUN; QUIT;
```

Appendix 9.

Solutions to Exercises

1. The scatterplots of the data for the two levels 1 and 2 of Org give the impression that a quadratic function of N may be suitable to describe the relation. It is not clear whether this is also a suitable approach for the variant without organic fertilization.
2. Assuming a quadratic function and considering the fact that the design is a randomized complete block design, the coincidence test rejects the hypothesis of coincidence with a Type 1 error rate of 0.05

(F-value = (2552.716682-1962.3931)/(10-4)/5.826405 =16.886 and p value < 0.0001).

To obtain the intermediate results for the coincidence test we used PROC GLM with the following CLASS and MODEL statements:

```
CLASS org block;
MODEL yield = org block N N*org N*N N*N*org/s;
CLASS block;
MODEL yield = block N N*N /s;
```

Alternatively, a joint test for coincidence is possible using a contrast statement in PROC GLM:

```
CLASS block org;
MODEL yield=block org org*n org*N*N / SOLUTION;
CONTRAST 'coincidence' org 1 -1 0,
                       org 1 0 -1,
                       org*N 1 -1 0,
                       org*N 1 0 -1,
                       org*N*N 1 -1,
                       org*N*N 1 0 -1;
```

The estimates and tests for the intercepts averaged across the blocks can be obtained by the statements

```
ESTIMATE 'intercept Org=0' intercept 3 block 1 1 1 org 3 0 0/ DIVISOR=3;
ESTIMATE 'intercept Org=1' intercept 3 block 1 1 1 org 0 3 0/ DIVISOR=3;
ESTIMATE 'intercept Org=2' intercept 3 block 1 1 1 org 0 0 3/ DIVISOR=3;
```

Based on the common estimated residual variance s^2 = 5.8264, the results for all regression parameters are:

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|
			Org = 0		
Intercept	36.4467	1.3936	22	26.15	< 0.0001
b1	0.08258	0.05131	22	1.61	0.1218
b2	0.00013	0.000348	22	0.37	0.7115
			Org = 1		
Intercept	34.6407	1.3583	22	25.5	< 0.0001
b1	0.5711	0.04363	22	13.09	< 0.0001
b2	-0.00291	0.000279	22	-10.45	< 0.0001
			Org = 2		
Intercept	36.3637	1.3583	22	26.77	< 0.0001
b1	0.3741	0.04363	22	8.57	< 0.0001
b2	-0.00197	0.000279	22	-7.07	< 0.0001

All parameters of the two levels with organic fertilization (Org = 1 and 2) are significantly different from zero; the regression coefficients of the variant Org = 0 are not significantly different from zero. The reason for their non-significance may be that in this case a linear function is appropriate and the quadratic approach is overparametrized (the tests are based on the partial approach).

As a consequence, polynomial functions of different order need to be considered for the different levels of organic fertilization, a situation which we did not discuss in this chapter. The need for this somewhat more complicated analysis does not arise if for all three levels of Org the same function type (e.g., a linear or quadratic function) is appropriate. We demonstrate two approaches to take this in consideration. The first one separates the analysis in two parts: the analysis for Org = 0 and the joint analysis for Org = 1 and 2. The second approach does not separate the analysis for the three Org levels using a trick which has several advantages compared to the first approach.

3.1 Approach with separate analyses of Org = 0 and joint analysis of Org = 1 and 2

If the function type is not the same for all levels of a factor, the first idea may be to analyze the data separately for each level. If we would analyze the three organic variants separately and consider the block structure of the experiment, for each variant different block means would be estimated. The efficient analysis of the RCB design requires, however, that all variants are analysed jointly. Moreover, we know from the F-test that there are significant differences between blocks. To consider the same block effects for all variants, we estimate block effects based on a joint analysis of all treatments as block effect = (total mean – mean of all values of the corresponding block), subtract these estimates from all observed values and then analyze the corrected data without block effects separately for each function type (Org = 0 and Org =1 and 2). In doing so, we do not regress observed values but corrected values (observed values – block effects) on the N fertilization rates. Therefore the degrees of freedom of the residuals have to be corrected.

3.1.1 Regression for the variant without organic fertilization

Due to our supposition that a linear approach is better suited for Org = 0, we choose a sequential approach starting with the linear term followed by a quadratic. For a polynomial regression analysis with a linear and a quadratic term and nine observed values, the DF of the residuals would normally be N-3 = 6. To consider the correction of the observed values by the block effects (DF blocks = 2), we set the denominator DF to N-3-2 = 4. This can be achieved in PROC MIXED by

```
MODEL yield_corr= N N2 / df=4,4 HTYPE=1;
```

The corrected yield corresponds to the variable `yield _ corr`. The parameter estimates are the same as above in the joint analysis and $s^2 = 7.2204$. The sequential F-tests confirm our supposition that a linear function is better suited.

Effect	Numerator DF	Denominator DF	*F* Value	Pr > *F*
N	1	4	49.63	0.0021
N2	1	4	0.11	0.7534

Finally, the fitted linear regression function is the following (using $s^2 = 6.3057$):

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	36.3071	1.3971	5†	25.99	< 0.0001
b1	0.1012	0.01342	5	7.54	0.0007

†For the linear function, we set the denominator DF = N-2-2 = 5.

3. 1. 2 Regressions for the variants with organic fertilization

Again, we use the corrected values. The denominator DF have been set to N-6-2 = 24-6-2 = 16 because the model has 6 regression parameters. The residual variance is common for both variants and is estimated as $s^2 = 4.7144$. Again, the regression parameter estimates are the same as in the joint analysis of all three variants.

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|
		Org = 1			
intercept	34.6407	1.2218	16	28.35	< 0.0001
b1	0.5711	0.03924	16	14.55	< 0.0001
b2	-0.00291	0.000251	16	-11.62	< 0.0001
		Org = 2			
intercept	36.3637	1.2218	16	29.76	< 0.0001
b1	0.3741	0.03924	16	9.53	< 0.0001
b2	-0.00197	0.000251	16	-7.86	< 0.0001

The disadvantage of this approach with separation of the analysis by the function type is that the estimation of the residual variance is not based on all observed values and that degrees of freedom are lost for the tests. Therefore, we do not discuss further options under this approach (test of coincidence of parameters or variance heterogeneity for Org = 1 and 2) and recommend instead the following joint analysis of all three levels of Org.

3.2 Regressions based on a joint analysis for Org = 0, 1, and 2

We want to do a joint analysis fitting a linear regression for Org = 0 and a quadratic regression for Org = 1, 2. The trick to suppress the quadratic term for Org = 0 is based on the definition of an auxiliary variable denoted as *switch* (Piepho et al., 2006). *switch* is equal to 0 if Org = 0 and it is equal to 1 if Org = 1 or 2 so that the variable switch*org*N*N is equal to 0 if Org = 0 and it is equal to org*N*N if Org = 1 or 2.

With this new variable *switch*, we analyze the data with the following CLASS and MODEL statements in PROC GLM and test the coincidence:

```
CLASS block org;
MODEL yield=block org org*n switch*org*N*N / SOLUTION;
CONTRAST 'coincidence' org 1 -1 0,
                       org 1 0 -1,
                       org*N 1 -1 0,
                       org*N 1 0 -1,
                       switch*org*N*N 1 -1,
                       switch*org*N*N 1 0 -1;
```

Whether the three intercepts, the three linear terms, and the two quadratic terms (Org =1 and 2) coincide can be tested by

```
CONTRAST 'coincidence intercept' org 1 -1 0,
                                 org 1 0 -1;
CONTRAST 'coincidence linear'    org*N 1 -1 0,
                                 org*N 1 0 -1;
CONTRAST 'coincidence quadratic' switch*org*N*N 0 1 -1;
```

The results are the following:

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
joint coincidence	5	1337.95021	267.590042	47.71	< 0.0001
coincidence intercept	2	6.072477	3.036238	0.54	0.5892
coincidence linear	2	778.172377	389.086188	69.37	< 0.0001
coincidence quadratic	1	33.27615	33.27615	5.93	0.023

We see that the three intercepts coincide, whereas the linear and quadratic terms do not. The corresponding estimates and tests are shown in the table below. The intercepts have been calculated as averages across blocks and the tests are based on the common estimated residual variance $s^2 = 5.6086$:

Parameter	Estimate	Standard Error	DF	*t* Value	Pr > |*t*|
		Org = 0			
Intercept	36.3071	1.31758	23	27.56	< 0.0001
b1	0.1012	0.01266	23	7.99	< 0.0001
		Org = 1			
Intercept	34.6407	1.33269	23	25.99	< 0.0001
b1	0.5711	0.04280	23	13.34	< 0.0001
b2	-0.00291	0.000273	23	-10.65	< 0.0001
		Org = 2			
Intercept	36.3637	1.33269	23	27.29	< 0.0001
b1	0.3741	0.04280	23	8.74	< 0.0001
b2	-0.00197	0.000273	23	-7.21	< 0.0001

The estimated regression parameters are identical to those obtained by the first approach; the test results differ slightly due to differences in the residual variance used and differences in the DF.

The intercepts of the three levels of org do not differ significantly, so we fit a model with a common intercept. At first, we test with PROC MIXED whether variant-specific residual variances or a common one should be assumed for this model.

```
CLASS block org;
MODEL yield=block org*n switch*org*N*N / SOLUTION;
REPEATED/ GROUP = org;
```

The LR-test for the comparison of the two models indicates a better fit of the model with a common variance (p value = 0.1115). Therefore, we delete the repeated statement in PROC MIXED.

The common intercept in the mean of the blocks can be estimated and tested by the statement

```
ESTIMATE 'intercept ' intercept 3 block 1 1 1 /DIVISOR=3;
```

Finally, we obtain the following results assuming a common variance ($s^2 = 5.4028$):

Parameter	Estimate	Standard Error	DF	*t* Value	Pr > |*t*|
		Org = 0			
Intercept	35.7746	0.7523	25	47.55	< 0.0001
b1	0.1053	0.00943	25	11.17	< 0.0001
		Org = 1			
Intercept	35.7746	0.7523	25	47.55	< 0.0001
b1	0.5460	0.03472	25	15.73	< 0.0001
b2	-0.00279	0.00024	25	-11.47	< 0.0001
		Org = 2			
Intercept	35.7746	0.7523	25	47.55	< 0.0001
b1	0.3871	0.03472	25	11.15	< 0.0001
b2	-0.00203	0.00024	25	-8.35	< 0.0001

Reference

Piepho, H.P., E.R. Williams, and M. Fleck. 2006. A note on the analysis of designed experiments with complex treatment structure. HortScience 41:446–452.

CHAPTER 7: ANALYSIS AND INTERPRETATION OF INTERACTIONS OF FIXED AND RANDOM EFFECTS

Mateo Vargas, Barry Glaz, Jose Crossa, and Alex Morgounov

Appendix 1: SAS code for orthogonal polynomial contrasts and graphs

SAS code for calculating the orthogonal polynomial contrasts

The following SAS code can be used for calculating the orthogonal polynomial coefficients for any contrast of interest and applied for the analyses of the example shown in Section 1. This procedure generates the correct coefficients for levels or rates that are equally or unequally spaced (as for the P rates we used). It is necessary to have installed the Interactive Matrix Language (**IML**) procedure. Our example here will show only how to calculate coefficients for P, because this is our only factor with quantitative rates or levels > 2. We show the code for calculating the contrasts for the main effect of P, and exclusively one two-way (from three), and one three-way (from three), and the unique four-way interactions with P, the extension to the other interactions is straightforward. We are including the four-way interaction contrast, even though this was not included in the reduced model. In Appendix 2, we show the complete code for calculating exhaustively all the possible contrasts involving P in the four-way analysis of variance

SAS macro program

```
**** Reading data: Data Wheat is available in attached CSV file ****;
Data Wheat;
      Infile "C:\Experiment 1 Data Wheat.CSV" dlm="," firstobs=3;
      Informat Soil$ 10.;
      Input Year Soil N P Rep Yield ;
Datalines;
Run;
**** Here begins the macro code for calculating the orthogonal
polynomial coefficients in an automatic way ****;
%Macro Coefficients;
Proc IML;
Plevels = {0, 50, 150, 250};
Coeff = Orpol(Plevels,3);
Ncoef = ncol(coeff)-1;
Call symputx("ncoef", ncoef);
      %do K = 1%to &ncoef;
           CoefGrade&K = t(Coeff[,&K+1]);
           PosCoefGrade&K = rowcat(char( CoefGrade&K ,15,10));
           NegCoefGrade&K = rowcat(char(-(CoefGrade&K),15,10));
           Call symputx("PosCoefGrade&K", PosCoefGrade&K);
           Call symputx("NegCoefGrade&K", NegCoefGrade&K);
      %end;
Run;
Title1 "Four-way ANOVA, decomposing df for P into three contrasts";
Proc GLIMMIX Data = Wheat;
      Class Year Soil N P Rep;
      Model Yield = Year | Soil | N | P;
      Random Rep(Year Soil);
**** P main effects contrasts ****;
```

```
Contrast "Linear P" P &PosCoefGrade1 ;
Contrast "Quadratic P" P &PosCoefGrade2 ;
Contrast "Cubic P" P &PosCoefGrade3 ;
**** Year × P two-way interaction contrasts ****;
Contrast "Linear Y*P" Year*P &PosCoefGrade1 &NegCoefGrade1 ;
Contrast "Quadratic Y*P" Year*P &PosCoefGrade2 &NegCoefGrade2 ;
Contrast "Cubic Y*P" Year*P &PosCoefGrade3 &NegCoefGrade3 ;
**** Year × Soil × P three-way interaction contrasts ****;
Contrast "Linear Y*S*P" Year*Soil*P
      &PosCoefGrade1 &NegCoefGrade1 &NegCoefGrade1 &PosCoefGrade1;
Contrast "Quadratic Y*S*P" Year*Soil*P
      &PosCoefGrade2 &NegCoefGrade2 &NegCoefGrade2 &PosCoefGrade2;
Contrast "Cubic Y*S*P" Year*Soil*P
      &PosCoefGrade3 &NegCoefGrade3 &NegCoefGrade3 &PosCoefGrade3;
**** Year × Soil × N × P four-way interaction contrasts ****;
Contrast "Linear Y*S*N*P" Year*Soil*N*P
      &PosCoefGrade1 &NegCoefGrade1 &NegCoefGrade1 &PosCoefGrade1
      &NegCoefGrade1 &PosCoefGrade1 &PosCoefGrade1 &NegCoefGrade1;
Contrast "Quadratic Y*S*N*P" Year*Soil*N*P
      &PosCoefGrade2 &NegCoefGrade2 &NegCoefGrade2 &PosCoefGrade2
      &NegCoefGrade2 &PosCoefGrade2 &PosCoefGrade2 &NegCoefGrade2;
Contrast "Cubic Y*S*N*P" Year*Soil*N*P
      &PosCoefGrade3 &NegCoefGrade3 &NegCoefGrade3 &PosCoefGrade3

      &NegCoefGrade3 &PosCoefGrade3 &PosCoefGrade3 &NegCoefGrade3;
Run;
%Mend;
      %Coefficients;
Run;
```

Brief explanation of preceding SAS code

In the **Plevels** statement it is necessary to express how many and which levels of the variable you wish to calculate. For instance, for the data set we used in Section 1 of this chapter, P has 4 unevenly spaced rates: 0, 50, 150, and 250 kg ha^{-1}. The **Orpol** function is used for obtaining the required coefficients; since there are three degrees of freedom, we can calculate three polynomial coefficients––linear, quadratic, and cubic––which are assigned to macro variables that are later used in the data step in the complete four-way ANOVA and contrasts.

In the contrast statements for the main effects, we need only positive values of the coefficients, but in the two-, three, and four-way contrasts, a combination of positive and negative coefficients are required, depending on the levels of each of the factors involved in those interactions. In Appendix 2, we show the procedure for obtaining the corresponding order of signs for each contrast. These positive and negative coefficients are assigned to the **PosCoefGrade** and **NegCoefGrade** macro variables, respectively, using the **Call Symput** function. The order of the polynomial contrasts is obtained using the **%do** - **%end** cycle. Note that the appropriate number and grade of the coefficients are automatically determined in the **ncoef = ncol(coeff)-1** statement. This program can be easily modified for different numbers and/or levels of factors.

It is important to mention that the order of the interaction coefficients in the CONTRAST statement depends on the order of the factors that are listed in the CLASS statement. If this is not the case the program may be executed without any errors but you may not get the intended contrast. This will be shown in a more detailed manner in the Exercise 1 of the Appendix 3.

Once we have the appropriate coefficients, the contrasts can be calculated using SAS procedures such as GLM, MIXED, or GLIMMIX.

SAS Code for Graphing Interactions

Graphing the two-way interaction Year × P

The following code is useful for graphing a two-way interaction, the generalization to a three-way or four-way interaction is straightforward as will be shown in Appendix 2.

```
**** Reading data: Data Wheat is available in attached CSV file ****;
Data Wheat;
      Infile "C:\Experiment 1 Data Wheat.CSV" dlm="," firstobs=3 ;
      Informat Soil$ 10. ;
      Input Year Soil N P Rep Yield ;
Datalines;
Run;
ODS select Covparms Tests3 Lsmeans LSMLines Meanplot; Title1 " ";
Title2 "Four-way ANOVA and LSD for Year x P Interaction using the
Confidence Interval" ;
Proc GLIMMIX Data = Wheat ;
      Class Year Soil N P Rep ;
      Model Yield = Year Soil N P
                    Year*Soil Year*N Year*P Soil*N Soil*P
                    Year*Soil*N Year*Soil*P Soil*N*P ;
                    Random Rep(Year Soil);
      LSMeans Year*P / Lines CL Plot = mean (sliceby = Year Join CL);
      ODS OUTPUT LSMeans = LSMeans ;
Run;
***** Generating the different curves to be used in graphing the LSD bars *****;
Data Graph;
      Set LSMeans ;
      HWCI=Estimate-Lower;
      LSD = Sqrt(2)*(HWCI);
      if year = 2007 then Y_07=Estimate;
      if year = 2008 then Y_08=Estimate;
      if year = 2007 then do;
           Yield1=Estimate; output;
           Yield1=Estimate - LSD; output;
           Yield1=Estimate + LSD; output;
      end;
      if year = 2008 then do;
Yield2=Estimate; output;
Yield2=Estimate -LSD;
output;
Yield2=Estimate +LSD;
output;
end;
Proc GPlot Data = Graph ;
      Plot (Y_07 Y_08)*P (Yield1 Yield2)*P / frame overlay
                          vaxis = axis1 haxis = axis2 nolegend;
Symbol1 v=dot cv=black h = 2.0 l=1 w=2 i=rq ci=black ;
Symbol2 v=dot    cv=red   h = 2.0  l=1 w=2  i=RQ        ci=Red ;
Symbol3                            l=1 w=2  i=hiloct    ci=black ;
Symbol4                            l=1 w=2  i=hiloct    ci=red ;
axis1 length = 4.5 in order = (0.8 to 2.4 by 0.4)
label=(f=Albany h=2.0 a=90 r=0 "Grain yield (Mg ha-1) ")
      value=(f=Albany
      h=2.0 ) offset = (1) minor=none;
axis2 length = 4.5 in order = (0 to 250 by 50)
      label=(f=Albany h=2.0 "P fertilizer rate (kg ha-1) ")
      value=(f=Albany h=2.0 ) offset = (3) minor=none;
Run;
```

Brief explanation of preceding SAS code

The output delivery system (ODS), ***ODS select Covparms Tests3 LSMLines Meanplot*** is for saving exclusively the useful information. ***ODS Output LSMeans = LSMeans*** statement is useful for creating a temporary file with only the information that will be needed later. For generating the different variables containing information for the regression lines associated with each combination of the Year × P interaction, we created the new variables **Y_07** and **Y_08** from the yield values. Similarly, cycles **if – then do – end** are used for obtaining the information needed in the upper and lower LSD bars. The codes can be adapted if, for example, four regression lines need to be depicted with their corresponding LSD error bars, as in the Soil × N × P interaction shown in Fig. 1.3 from Section 1 and as provided with code in Example 2 of Appendix 2, and so on to any number of regression lines needed.

If we are interested in the ANOVA for only the terms found to be significant in the final model (Table 1.2), i.e. the four main effects, five two-way, and three three-way interactions, without including the four-way interaction; all those terms should be included in the model statement when computing the correct LSD and/or the confidence interval. In the LSMeans statement, we have included only the interaction that is of interest for graphing the LSD values, thus simplifying the output.

The selected output includes only three sections: with the covariance parameters estimates and the Type III test of fixed effects we have obtained all the information shown in Table 1.2.

The GLIMMIX Procedure

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error
Rep(Year*Soil)	0.000222	0.000953
Residual	0.008593	0.002084

Type III Tests of Fixed Effects

Effect	Num DF	**Den DF**	***F* Value**	**Pr > *F***
Year	1	4	1204.82	< 0.0001
Soil	1	4	556.51	< 0.0001
N	1	34	11.36	0.0019
P	3	34	37.75	< 0.0001
Year*Soil	1	4	125.30	0.0004
Year*N	1	34	9.45	0.0041
Year*P	3	34	6.40	0.0015
Soil*N	1	34	19.80	< 0.0001
Soil*P	3	34	13.72	< 0.0001
Year*Soil*N	1	34	33.31	< 0.0001
Year*Soil*P	3	34	7.94	0.0004
Soil*N*P	6	34	3.21	0.0131

From the Year*P Least Squares Means (lsmeans) table we can calculate the LSD value considering the confidence interval given by the upper and lower values, and then multiplying the half width of the interval (upper value minus the estimate or estimate minus lower value) by the square root of 2, as shown in the SAS code in the statements ***HWCI=Estimate-Lower; LSD = Sqrt(2)*(HWCI)***.

The estimates for the least squares means are then used for graphing the response curves by means of the ***GPLOT*** Procedure and using an interpolation method; the highest degree of the orthogonal polynomial contrast that was found to be significant in that particular two-way interaction, in this case for Year × P interaction in Table 1.2 was the quadratic contrast. This was performed with the code ***i=rq***, were ***i*** means interpolation and ***rq*** means Regression Quadratic, thus using the following statements and obtaining the graph shown below.

```
Symbol1 v=dot cv=black h = 2.0 l=1 w=2 i=rq ci=black ;
Symbol2 v=dot cv=red h = 2.0 l=1 w=2 i=rq ci=red ;
```

Year*P Least Squares Means

Year	P	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
2007	0	1.6537	0.03361	34	49.20	< 0.0001	0.05	1.5854	1.7221
2007	50	1.8600	0.03361	34	55.34	< 0.0001	0.05	1.7917	1.9283
2007	150	2.0537	0.03361	34	61.10	< 0.0001	0.05	1.9854	2.1221
2007	250	2.0838	0.03361	34	62.00	< 0.0001	0.05	2.0154	2.1521
2008	0	0.9138	0.03361	34	27.19	< 0.0001	0.05	0.8454	0.9821
2008	50	1.0213	0.03361	34	30.38	< 0.0001	0.05	0.9529	1.0896
2008	150	1.0563	0.03361	34	31.43	< 0.0001	0.05	0.9879	1.1246
2008	250	1.1250	0.03361	34	33.47	< 0.0001	0.05	1.0567	1.1933

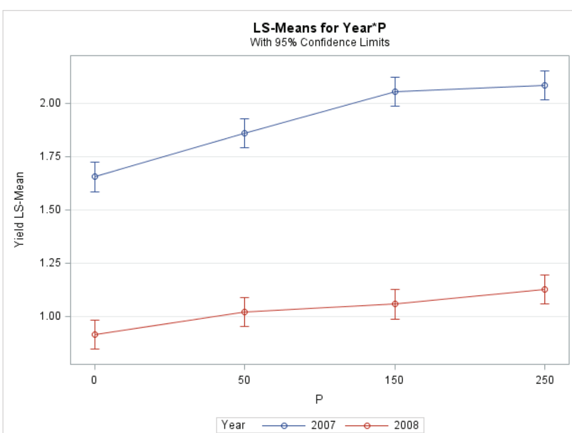
The next section of output includes the LSD grouping of the lsmeans which are calculated using the option ***Lines*** in the ***LSMeans*** statement as follows:

```
LSMeans Year*P / Lines CL Plot = mean (sliceby = Year Join CL);
```

T Grouping for Year*P
Least Squares Means (Alpha=0.05)
LS-means with the same letter are not significantly different.

Year	P	Estimate		
2007	250	2.0838		A
2007	150	2.0537		A
2007	50	1.8600		B
2007	0	1.6537		C
2008	250	1.1250		D
2008	150	1.0563	E	D
2008	50	1.0213	E	
2008	0	0.9138		F

Finally, the option ***Plot = mean (sliceby = Year Join CL)***, directly creates the following graph of means responses for each year (sliceby=year), joining the mean values using a straight line (Join) and including the 95% confidence limits (CL).



As you can see these graphs are directly and easily obtained using a single statement. However, the main limitation for this kind of graph is that you can only use straight lines as response curves and you cannot use a more sophisticated interpolation method as you can using the ***GPLOT*** procedure.

In Appendix 2, we show other examples for obtaining more complicated graphs, for three- and four-way interactions, and we will show that the extension from this two-way interaction case is straightforward.

Appendix 2: Generalization of SAS code programs shown in Appendix 1

Example 1: SAS macro code for calculating orthogonal polynomial coefficients

In this Appendix, we describe the complete SAS code for calculating all possible orthogonal polynomial contrasts involving P in the data set Wheat. This will include the contrasts for the main effect of P, the six two-way interactions, the four three-way interactions, and the four- way interaction. Although we recommended not including the four-way interaction in our reduced analysis in Section 1, as a resource for readers, we will include here the code for the contrasts associated with this interaction.

As stated in Appendix 1, it is necessary to have installed the Interactive Matrix Language (**IML**) procedure for calculating the orthogonal polynomial coefficients. In the contrast statements for the main effects, we need only positive values of the coefficients, but in the two-, three, and four-way contrasts, positive and negative coefficients are required depending on the levels of each of the factors involved in those interactions. For obtaining the order of signs of the coefficients in the corresponding contrast, one can use Table A2.1 (in this appendix) as an example. This table is only for the linear contrasts (CoefGrade1) where CoefGrade1 is a vector of length 1 × 4 (1 row, 4 columns), one coefficient value for each P rate.

One may use the coefficient -1 for Year 1 and the coefficient +1 for Year 2, and so on, for any factor: -1 for Black soil and +1 for Chesnut soil; -1 for the first rate of N (N0) and +1 for the second rate of N (N30). In fact, this selection of signs is what is used in a factorial experiment 2^k, where 2 is the number of levels of each factor and *k* is the number of factors. The order of positive and negative coefficients is not important because the significance of each contrast is based on sum of squares of differences among the levels tested. Therefore, the sign in these differences is cancelled by squaring them. However, if you are using the *Estimate* statement instead of the *Contrast* statement, the sign is important, because the interpretation would reverse. The significance of the comparison does not differ between the *Contrast* and *Estimate* statements. That is, the standard errors, the t values, and the p-values are exactly the same despite the choice of which level to consider as -1 or +1. Again, the significance of a *Contrast* or an *Estimate* statement, whichever you prefer to use, will be equivalent.

Table A2.1. Coefficients for the linear contrast for P main effect, two-way, three-way, and four-way interactions.

Effect	Factor 1	Coefficient	Factor 2	Coefficient	Factor 3	Coefficient	Phosphorus fertilizer rates
P main effect	P	1					PosCoefGrade1
Two-way Y × P Interaction	Year 1	1					PosCoefGrade1
	Year 2	-1					NegCoefGrade1
Three-way Y × Soil × P Interaction	Year 1	1	Black	1			PosCoefGrade1
			Chestnut	-1			NegCoefGrade1
	Year 2	-1	Black	1			NegCoefGrade1
			Chestnut	-1			PosCoefGrade1
Four-way Y × Soil × N × P Interaction	Year 1	1	Black	1	N_0	1	PosCoefGrade1
					N_{30}	-1	NegCoefGrade1
			Chestnut	-1	N_0	1	NegCoefGrade1
					N_{30}	-1	PosCoefGrade1
	Year 2	-1	Black	1	N_0	1	NegCoefGrade1
					N_{30}	-1	PosCoefGrade1
			Chestnut	-1	N_0	1	PosCoefGrade1
					N_{30}	-1	NegCoefGrade1

Therefore, using the sequence of signs shown in Table A2.1, the complete SAS code for estimating all the possible polynomial contrasts involving the P factor, is the following:

SAS macro program

```
**** Reading data: Data Wheat is available in attached CSV file ****;
Data Wheat;
     Infile "C:\Experiment 1 Data Wheat.CSV" dlm="," firstobs=3 ;
     Informat Soil$ 10. ;
     Input Year Soil N P Rep Yield ;
Datalines;
Run;
**** Here begins the macro code for calculating the orthogonal
polynomial contrasts in an automatic way ****;
%Macro Coefficients;
Proc IML;
     Plevels = {0, 50, 150, 250};
     Coeff = Orpol(Plevels,3);
     Ncoef = ncol(coeff)-1;
     Call symputx("ncoef", ncoef);
     %do K = 1%to &ncoef;
          CoefGrade&K = t(Coeff[,&K+1]);
          PosCoefGrade&K = rowcat(char( CoefGrade&K ,15,10));
          NegCoefGrade&K = rowcat(char(-(CoefGrade&K),15,10));
          Call symputx("PosCoefGrade&K", PosCoefGrade&K);
          Call symputx("NegCoefGrade&K", NegCoefGrade&K);
     %end;
Run;
Title1 "Four-way ANOVA, decomposing df for P into three contrasts";
Proc GLIMMIX Data = Wheat;
     Class Year Soil N P Rep;
     Model Yield = Year | Soil | N | P;
     Random Rep(Year Soil);
**** P main effects contrasts ****;
Contrast "P Linear " P &PosCoefGrade1 ;
Contrast "P Quadratic " P &PosCoefGrade2 ;
Contrast "P Cubic" P &PosCoefGrade3 ;
**** Year × P two-way interaction contrasts ****;
Contrast "Y × P Linear " Year*P &PosCoefGrade1 &NegCoefGrade1 ;
Contrast "Y × P Quadratic" Year*P &PosCoefGrade2 &NegCoefGrade2 ;
```

```
 Contrast "Y × P Cubic" Year*P &PosCoefGrade3 &NegCoefGrade3 ;
**** Soil × P two-way interaction contrasts ****;
 Contrast "S × P Linear" Soil*P &PosCoefGrade1 &NegCoefGrade1 ;
 Contrast "S × P Quadratic" Soil*P &PosCoefGrade2 &NegCoefGrade2 ;
 Contrast "S × P Cubic" Soil*P &PosCoefGrade3 &NegCoefGrade3 ;
**** N × P two-way interaction contrasts ****;
 Contrast "N × P Linear" N*P &PosCoefGrade1 &NegCoefGrade1 ;
 Contrast "N × P Quadratic" N*P &PosCoefGrade2 &NegCoefGrade2 ;
 Contrast "N × P Cubic " N*P &PosCoefGrade3 &NegCoefGrade3 ;
**** Year × Soil × P three-way interaction contrasts ****;
 Contrast "Y × S × P Linear" Year*Soil*P
      &PosCoefGrade1 &NegCoefGrade1 &NegCoefGrade1 &PosCoefGrade1 ;
 Contrast "Y × S × P Quadratic" Year*Soil*P
      &PosCoefGrade2 &NegCoefGrade2 &NegCoefGrade2 &PosCoefGrade2 ;
 Contrast "Y × S × P Cubic" Year*Soil*P
      &PosCoefGrade3 &NegCoefGrade3 &NegCoefGrade3 &PosCoefGrade3 ;
**** Year × N × P three-way interaction contrasts ****;
 Contrast "Y × N × P Linear" Year*N*P
      &PosCoefGrade1 &NegCoefGrade1 &NegCoefGrade1 &PosCoefGrade1 ;
 Contrast "Y × N × P Quadratic" Year*N*P
      &PosCoefGrade2 &NegCoefGrade2 &NegCoefGrade2 &PosCoefGrade2 ;
 Contrast "Y × N × P Cubic" Year*N*P
      &PosCoefGrade3 &NegCoefGrade3 &NegCoefGrade3 &PosCoefGrade3 ;
**** Soil × N × P three-way interaction contrasts ****;
 Contrast "S × N × P Linear" Soil*N*P
      &PosCoefGrade1 &NegCoefGrade1 &NegCoefGrade1 &PosCoefGrade1 ;
 Contrast "S × N × P Quadratic" Soil*N*P
      &PosCoefGrade2 &NegCoefGrade2 &NegCoefGrade2 &PosCoefGrade2 ;
 Contrast "S × N × P Cubic" Soil*N*P
      &PosCoefGrade3 &NegCoefGrade3 &NegCoefGrade3 &PosCoefGrade3 ;
**** Year × Soil × N × P four-way interaction contrasts ****;
 Contrast "Y × S × N × P Linear" Year*Soil*N*P
      &PosCoefGrade1 &NegCoefGrade1 &NegCoefGrade1 &PosCoefGrade1
      &NegCoefGrade1 &PosCoefGrade1 &PosCoefGrade1 &NegCoefGrade1;
 Contrast "Y × S × N × P Quadratic" Year*Soil*N*P
      &PosCoefGrade2 &NegCoefGrade2 &NegCoefGrade2 &PosCoefGrade2
      &NegCoefGrade2 &PosCoefGrade2 &PosCoefGrade2 &NegCoefGrade2;
 Contrast "Y × S × N × P Cubic" Year*Soil*N*P
      &PosCoefGrade3 &NegCoefGrade3 &NegCoefGrade3 &PosCoefGrade3
      &NegCoefGrade3 &PosCoefGrade3 &PosCoefGrade3 &NegCoefGrade3;
 Run;
 %Mend;
 %Coefficients;
 Run;
```

This SAS code was discussed previously in Appendix 1. Here, we are only adding all possible contrasts.

The selected output for this complete four-way model

The covariance parameter estimates and the Type III tests of fixed effects were already shown in Table 1.1 of Section 1. We repeat that information here and we also show here the complete results for all the possible contrasts involving the P rates from the Experiment 1 data.

The GLIMMIX Procedure

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
Rep(Year*Soil)	0.000292	0.000955
Residual	0.008036	0.002148

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Year	1	4	1204.82	< 0.0001

Soil	1	4	556.51	< 0.0001
N	1	28	12.15	0.0016
P	3	28	40.37	< 0.0001
Year*Soil	1	4	125.30	0.0004
Year*N	1	28	10.11	0.0036
Year*P	3	28	6.84	0.0013
Soil*N	1	28	21.17	< 0.0001
Soil*P	3	28	14.67	< 0.0001
N*P	3	28	1.49	0.2400
Year*Soil*N	1	28	35.62	< 0.0001
Year*Soil*P	3	28	8.50	0.0004
Year*N*P	3	28	0.65	0.5892
Soil*N*P	3	28	5.39	0.0047
Year*Soil*N*P	3	28	2.13	0.1183

Contrasts				
Label	Num DF	Den DF	*F* Value	Pr > *F*
P Linear	1	28	106.69	< 0.0001
P Quadratic	1	28	12.60	0.0014
P Cubic	1	28	1.82	0.1886
Y × P Linear	1	28	15.21	0.0006
Y × P Quadratic	1	28	5.09	0.0321
Y × P Cubic	1	28	0.24	0.6285
S × P Linear	1	28	43.81	< 0.0001
S × P Quadratic	1	28	0.19	0.6635
S × P Cubic	1	28	0.00	0.9801
N × P Linear	1	28	2.59	0.1190
N × P Quadratic	1	28	0.48	0.4927
N × P Cubic	1	28	1.39	0.2491
Y × S × P Linear	1	28	24.96	< 0.0001
Y × S × P Quadratic	1	28	0.02	0.8787
Y × S × P Cubic	1	28	0.50	0.4835
Y × N × P Linear	1	28	0.82	0.3718
Y × N × P Quadratic	1	28	0.49	0.4902
Y × N × P Cubic	1	28	0.64	0.4307
S × N × P Linear	1	28	3.00	0.0943
S × N × P Quadratic	1	28	0.07	0.7866
S × N × P Cubic	1	28	13.09	0.0012
Y × S × N × P Linear	1	28	0.50	0.4870
Y × S × N × P Quadratic	1	28	0.01	0.9436
Y × S × N × P Cubic	1	28	5.90	0.0218

Generalization of SAS Code for Graphing Interactions

Example 2: Graphing the three-way interaction Year × Soil × P

Here we present an example for graphing a three-way interaction. For Fig. 1.2, we are interested in the ANOVA for all the terms which were significant, that is, the four main effects, five two-way, and three three-way interactions; thus all those terms should be included in the model statement when computing the correct LSD and/or the confidence interval. However, in the LSMeans statement, we have included only the interactions that are of interest for graphing the LSD values in order to simplify the output.

```
**** Reading data: Data Wheat is available in attached CSV file ****;
Data Wheat;
Infile "C:\Experiment 1 Data Wheat.CSV" dlm="," firstobs=3 ;
```

```
Informat Soil$ 10. ;
Input Year Soil N P Rep Yield ;
Datalines;
Run;
ODS select CovParms Tests3 LSMLines Meanplot;
Title2 "Four-way ANOVA and LSD for Year x Soil x P Interaction using
the Confidence Interval";
Proc GLIMMIX data = Wheat;
      Class Year Soil N P Rep ;
      Model Yield = Year Soil N P
                     Year*Soil Year*N Year*P Soil*N Soil*P
                     Year*Soil*N Year*Soil*P Soil*N*P;
      Random Rep(Year Soil);
LSMeans Year*Soil*P / Lines CL plot=mean(sliceby=Year*Soil Join CL);
ODS Output LSMeans = LSMeans;
Run;
***** Generating the different curves to be used in graphing the LSD
bars *****;
Data Graph;
      Set LSMeans ;
      HWCI = Estimate - Lower;
      LSD = Sqrt(2)*(HWCI);
if Year = 2007 and Soil = "black" then Y1S1 = Estimate ;
if Year = 2007 and Soil = "chestnut" then Y1S2 = Estimate ;
if Year = 2008 and Soil = "black" then Y2S1 = Estimate ;
if Year = 2008 and Soil = "chestnut" then Y2S2 = Estimate ;
if Year = 2007 and Soil = "black" then do;
      Yield1 = Estimate; output;
      Yield1 = Estimate - (LSD/2); output;
      Yield1 = Estimate + (LSD/2); output;
end;
if Year = 2007 and Soil = "chestnut" then do;
      Yield2 = Estimate; output;
      Yield2 = Estimate - (LSD/2); output;
      Yield2 = Estimate + (LSD/2); output;
end;
if Year = 2008 and Soil = "black" then do;
      Yield3 = Estimate; output;
      Yield3 = Estimate - (LSD/2); output;
      Yield3 = Estimate + (LSD/2); output;
end;
if Year = 2008 and Soil = "chestnut" then do;
      Yield4 = Estimate; output;
      Yield4 = Estimate - (LSD/2); output;
      Yield4 = Estimate + (LSD/2); output;
end;
Run;
** Graphics options for creating an enhanced meta-file **;
FILENAME Figure 'C:\Output\Figure 2.1, Y x S x P.EMF';
GOPTIONS DEVICE=SASEMF GSFNAME=Figure GSFMODE=Replace;
Proc Gplot data = Graph ;
      Plot (Y1S1 Y1S2 Y2S1 Y2S2)*P (Yield1 Yield2 Yield3 Yield4)*P /
           overlay frame vaxis = axis1 haxis = axis2 nolegend;
Symbol1 v=dot cv=black h = 2.0 l=1 w=2 i=rl ci=black ;
Symbol2 v=dot cv=blue h = 2.0 l=1 w=2 i=rq ci=blue ;
Symbol3 v=dot cv=green h = 2.0 l=1 w=2 i=rl ci=green ;
Symbol4 v=dot cv=red h = 2.0 l=1 w=2 i=rl ci=Red ;
Symbol5 l=1 w=2 i=hiloct ci=black;
Symbol6 l=1 w=2 i=hiloct ci=blue;
Symbol7 l=1 w=2 i=hiloct ci=green;
Symbol8 l=1 w=2 i=hiloct ci=Red;
axis1 length = 4.5 in order = (0.5 to 3.0 by 0.5)
      label=(f=Albany h=3.0 a=90 r=0 "Grain yield (Mg ha-1)")
      value=(f=Albany h=3.0) offset = (1) minor=none;
axis2 length = 7.0 in order = (0 to 250 by 50)
      label=(f=Albany h=3.0 "P fertilizer rate (kg ha-1)")
      value=(f=Albany h=3.0) offset = (3) minor=none;
Title1 f=Albany h=2.0 "Figure 1.2.- Year × Soil × P Interaction
response profiles";
Run;
```

Brief Description of SAS code

The **ODS select *CovParms Tests3 LSMLines Meanplot*** option is used to save the useful and needed output: the covariance parameters of estimates for random terms (***CovParms***); the Type III tests of fixed effects (***Tests3***); the least squares means and their t (LSD) grouping (***LSMLines***); and finally the mean response profiles for the three-way interaction (***Meanplot***). The SAS system is insensitive to small or capital letters, we use a mix of both only for emphasis.

The ***meanplot*** option or simply ***mean*** requests that the least squares means (lsmeans) be displayed. For example, in the line ***LSMeans Year*Soil*P / Lines CL plot=mean(sliceby=Year*Soil Join CL)***, the lsmeans response profiles are requested for the three- way Year × Soil × P interaction. The ***meanplot-options*** controls the display of the least square means; ***join*** or ***connect*** connects the lsmeans with lines; ***Sliceby=Year*Soil*** creates four response profiles coming from the two-way interaction Year × Soil at each P rate; and the ***CL*** code displays upper and lower confidence limits for the lsmeans. By default, 95% limits are drawn. The confidence levels can be changed with the **alpha**= option. In the next example, we will ask for the lsmeans, LSD grouping, and the response profiles for all the main effects, two-way, three-way, and the four-way interactions.

The second part of the SAS program is for calculating and graphing the response profiles for each mean using the ***GPLOT*** procedure, in order to include a different interpolation curve for each profile, rather than simply using lines joined in the default output of the ***GLIMMIX*** procedure.

Using information saved in the temporary file named ***LSMeans*** in the ODS Output ***LSMeans = LSMeans*** statement, we calculate first the half width of the confidence interval using the ***HWCI = Estimate – Lower*** statement, and finally we calculate the LSD value using the expression ***LSD = Sqrt(2)*(HWCI).***

In the next block of statements, using the first statement as an example: ***if Year = 2007 and Soil = "black" then Y1S1 = Estimate*** creates one response profile for the first Year × Soil combination (Year 1, Soil 1) and assigns the values to a new variable ***Y1S1,*** which will be used for graphing the corresponding response profiles, and so on for the other three Year × Soil combinations. Similarly the following ***if-then do-end*** cycle:

```
if Year = 2007 and Soil = "black" then do;
Yield1 = Estimate; output;
Yield1 = Estimate -(LSD/2); output;
Yield1 = Estimate + (LSD/2); output;
end;
```

is used for calculating the center, lower, and upper LSD bars to be included for each response profile at each P rate. This is similar for the other three Year × Soil combinations.

In the ***Plot (Y1S1 Y1S2 Y2S1 Y2S2)*P (Yield1 Yield2 Yield3 Yield4)*P / overlay*** statement for the ***GPLOT*** procedure we ask to simultaneously plot the four response profiles and the four LSD bars associated with each curve. Then, with the ***Symbol*** option, we can use a different interpolation method for each curve, using, for example, ***i=rl***, ***i=rq***, or ***i=rc***, for a linear, quadratic or cubic regression, respectively. For drawing the LSD bars, we used the interpolation alternative ***i=hiloct***. The generalization to a four-way interaction in which we need to include eight response profiles is straightforward from this example.

Complete Selected Output from this example

The results for the covariance parameter estimates and the Type III tests of fixed effects are the same here as those shown previously in Table 1.2 from Section 1, which corresponded to the reduced final model.

The GLIMMIX Procedure

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error
Rep(Year*Soil)	0.000222	0.000953
Residual	0.008593	0.002084

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Year	1	4	1204.82	< 0.0001
Soil	1	4	556.51	< 0.0001
N	1	34	11.36	0.0019
P	3	34	37.75	< 0.0001
Year*Soil	1	4	125.30	0.0004
Year*N	1	34	9.45	0.0041
Year*P	3	34	6.40	0.0015
Soil*N	1	34	19.80	< 0.0001
Soil*P	3	34	13.72	< 0.0001
Year*Soil*N	1	34	33.31	< 0.0001
Year*Soil*P	3	34	7.94	0.0004
Soil*N*P	6	34	3.21	0.0131

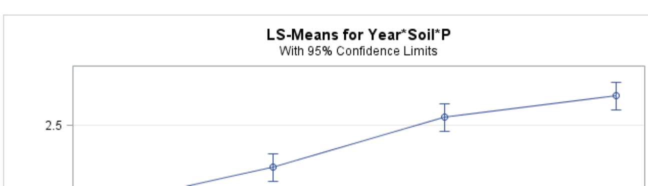
The following lsmeans and their LSD grouping were also shown in Table 1.4 of Section 1.

T Grouping for Year*Soil*P Least

Squares Means (Alpha=0.05)
LS-means with the same letter are not significantly different.

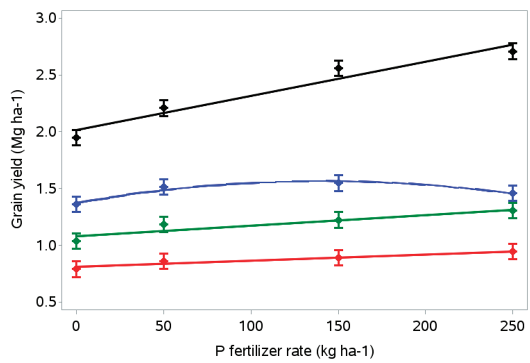
Soil	Year	P	Estimate		
black	2007	250	2.7075		A
black	2007	150	2.5600		B
black	2007	50	2.2075		C
black	2007	0	1.9475		D
chestnut	2007	150	1.5475		E
chestnut	2007	50	1.5125		E
chestnut	2007	250	1.4600	F	E
chestnut	2007	0	1.3600	F	G
black	2008	250	1.3050	H	G
black	2008	150	1.2225	H	
black	2008	50	1.1825	H	
black	2008	0	1.0375		I
chestnut	2008	250	0.9450	J	I
chestnut	2008	150	0.8900	J	K
chestnut	2008	50	0.8600	J	K
chestnut	2008	0	0.7900		K

The following graph showing the four response profiles for the Year × Soil × P interaction was obtained with the statement ***LSMeans Year*Soil*P / Lines CL plot=mean(sliceby =Year*Soil Join CL),*** already explained above.



Finally, the next plot was obtained using the ***GPLOT*** procedure and corresponds to Fig. 1.2 shown previously in Section 1 when explaining the significant interactions found in the final model.

Figure 1.2.- Year x Soil x P



Example 3: Four-way ANOVA, Mean Plots, and LSD values for Main Effects and all Two-, Three-, and Four-Way Interactions.

```
**** Reading data: Data Wheat is available in attached CSV file ****;
Data Wheat;
      Infile "C:\Experiment 1 Data Wheat.CSV" dlm="," firstobs=3 ;
      Informat Soil$ 10. ;
      Input Year Soil N P Rep Yield ; Datalines;
Run;
ODS select CovParms Tests3 Lsmeans Diffs LSMLines Meanplot;
Title2 "Four-way ANOVA and mean plots for all main effects and all
interactions";
Proc GLIMMIX data = Wheat;
      Class Year Soil N P Rep;
      Model Yield = Year | Soil | N | P ;
      Random Rep(Year Soil);
      LSMeans Year | Soil | N | P / PDIFF Lines CL;
      LSMeans Year Soil N P / plot = mean (Join CL);
      LSMeans Year*Soil Year*N Year*P/plot=mean (sliceby=Year Join CL);
      LSMeans Soil*N Soil*P / plot = mean (sliceby = Soil Join CL);
      LSMeans N*P / plot = mean (sliceby = N Join CL);
      LSMeans Year*Soil*N Year*Soil*P/plot=mean (sliceby=Year*Soil Join CL);
      LSMeans Year*N*P / plot = mean (sliceby = Year*N Join CL);
      LSMeans Soil*N*P / plot = mean (sliceby = Soil*N Join CL);
      LSMeans Year*Soil*N*P / plot=mean (sliceby = Year*Soil*N Join CL);
      ODS output lsmeans = LSMEANS diffs=DIFFS tests3=DOF;
Run;
*** Strategy for Calculating the LSD using the Average Standard Error
of Differences ***;
Proc Sort data = DOF;
      By Effect;
Proc Sort data = DIFFS;
      By Effect;
Proc Means data=DIFFS mean noprint;
      By Effect;
      Output out = ASED mean = AvStdErr;
      Var StdErr;
Data LSD;
      Merge ASED DOF;
      By Effect;
      t = tinv(1-0.05 / 2, DenDF);
      LSD = t*AvStdErr;
      Drop NumDF ;
```

```
Run;
If AvStdErr = . then delete ;
Title2 "LSD calculated using the Average Standard Error of Differences"
;
Proc Print Data = LSD ;
      Var Effect AvStdErr DenDF t LSD;
Run;
```

Brief description of the SAS code.

Almost all the components of this SAS code were explained previously in this appendix; now we will describe only additional statements used. Firstly, because we are interested in the complete model and in all lsmeans, we are using the bar notation of SAS in both the ***Model*** and ***LSMeans*** statements. The new option ***PDIFF*** in the ***LSMeans*** statement is used for generating the least squares differences, their standard errors, t values, and p-values, which will be saved in the temporary file ***DIFFS*** with the statement ***ODS output lsmeans diffs=DIFFS tests3=DOF***, and later used for calculating the LSD value by using the average standard error of differences (ASED) obtained in the related ***Proc Means*** block of statements**.**

The temporary file ***DOF*** is generated for saving the information related to the degrees of freedom (NumDF, DenDF) for numerator and denominator, respectively, in the Type III tests of fixed effects for each term included in the model. The ***DenDF*** are used later in asking SAS for the accumulated ***t*** probability using the ***t = tinv(1 - 0.05 / 2, DenDF)*** statement, and finally calculating the LSD values for each effect through the expression ***LSD = t*AvStdErr***.

In addition to the description of the statements used for graphing the response profiles in Example 2 above, now we will provide a more detailed explanation of this code. In the line ***LSMeans Year Soil N P / plot = mean (join cl)***, the lsmeans response profiles are requested for all the main effects Year, Soil, N, and P, simultaneously.

The statement ***LSMeans Year*Soil Year*N Year*P / plot = mean (sliceby = Year Join CL)*** is used for simultaneously requesting the response profiles for the Year × Soil, Year × N, and Year × P interactions. **Sliceby**=fixed-effect specifies the Year effect by which to group the means in a single plot, and the levels for the Soil, N, and P effects to be drawn in the horizontal axis, because Year is a qualitative factor while N and P are quantitative factors. Similarly, the statement ***LSMeans Soil*N Soil*P / plot = mean (sliceby = Soil Join CL)*** is used to draw the individual response profiles for each Soil at the N and P levels in the horizontal axis.

For separating means of three-way interactions, we can use a statement like the following: ***LSMeans Year*Soil*N Year*Soil*P / plot = mean (sliceby=Year*Soil Join CL)***, in which we are analyzing the three-way interactions Year × Soil × N and Year × Soil × P, creating four response profiles coming from the two-way interaction Year × Soil at each level of the N and P rates.

Finally, the statement ***LSMeans Year*Soil*N*P / plot = mean (sliceby = Year*Soil*N Join CL)*** is useful for drawing the eight response profiles for the three-way combination of the levels of factors Year, Soil, and N, all of them with two levels, and for each level of the P factor in the horizontal axis. Only one of the 15 graphs obtained with the provided SAS code are included in the SAS output (see Fig. 1.3, Section 1)

Partial Selected Output of SAS Code for Example 3

The covariance parameter estimates and Type III tests of fixed effects are the same as shown in the first example of this appendix. Instead of showing these again, we show the structure for the lsmeans, the differences, as well as the t grouping for the two-way Year × Soil interaction, as an example for visualizing the standard errors of the differences used for calculating the LSD values using the ASED.

Year*Soil Least Squares Means

Soil	Year	Estimate	Standard Error	DF	t Value	Pr >|t|	Alpha	Lower	Upper
black	2007	2.3556	0.02546	4	92.52	< 0.0001	0.05	2.2849	2.4263
chestnut	2007	1.4700	0.02546	4	57.74	< 0.0001	0.05	1.3993	1.5407
black	2008	1.1869	0.02546	4	46.62	< 0.0001	0.05	1.1162	1.2576
chestnut	2008	0.8712	0.02546	4	34.22	< 0.0001	0.05	0.8006	0.9419

Differences of Year*Soil Least Squares Means

Soil	Year	_Soil	_Year	Estimate	Standard Error	DF	t Value	Pr >|t|	Alpha	Lower	Upper
black	2007	chestnut	2007	0.8856	0.03601	4	24.60	< 0.0001	0.05	0.7857	0.9856
black	2007	black	2008	1.1687	0.03601	4	32.46	< 0.0001	0.05	1.0688	1.2687
black	2007	chestnut	2008	1.4844	0.03601	4	41.22	< 0.0001	0.05	1.3844	1.5843
chestnut	2007	black	2008	0.2831	0.03601	4	7.86	0.0014	0.05	0.1832	0.3831
chestnut	2007	chestnut	2008	0.5988	0.03601	4	16.63	< 0.0001	0.05	0.4988	0.6987
black	2008	chestnut	2008	0.3156	0.03601	4	8.77	0.0009	0.05	0.2157	0.4156

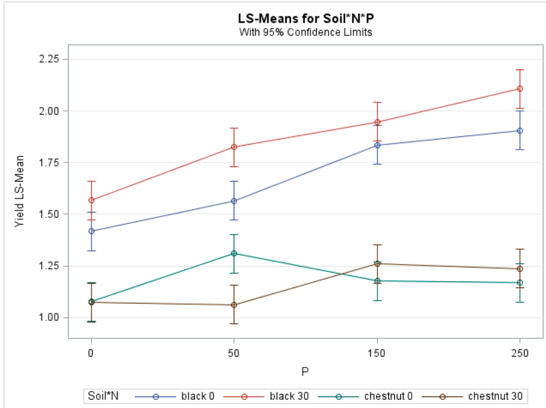
T Grouping for Year*Soil Least Squares Means (Alpha=0.05)
LS-means with the same letter are not significantly different.

Soil	Year	Estimate	
black	2007	2.3556	A
chestnut	2007	1.4700	B
black	2008	1.1869	C
chestnut	2008	0.8712	D

LSD calculated using the Average Standard Error of Differences

Obs	Effect	AvStdErr	DenDF	t	LSD
1	N	0.02241	28	2.04841	0.04591
2	N*P	0.04482	28	2.04841	0.09181
3	P	0.03169	28	2.04841	0.06492
4	Soil	0.02546	4	2.77645	0.07069
5	Soil*N	0.03318	28	2.04841	0.06796
6	Soil*N*P	0.06400	28	2.04841	0.13109
7	Soil*P	0.04574	28	2.04841	0.09369
8	Year	0.02546	4	2.77645	0.07069
9	Year*N	0.03318	28	2.04841	0.06796
10	Year*N*P	0.06400	28	2.04841	0.13109
11	Year*P	0.04574	28	2.04841	0.09369
12	Year*Soil	0.03601	4	2.77645	0.09997
13	Year*Soil*N	0.04752	28	2.04841	0.09734
14	Year*Soil*N*P	0.09089	28	2.04841	0.18619
15	Year*Soil*P	0.06520	28	2.04841	0.13355

Finally, following is the graph obtained for the three-way Soil × N × P interaction, like that shown in Fig. 1.3 from Section 1.



Appendix 3: Answers to Review Questions and Exercises

Review Questions: True or False

1. When the researcher can identify predetermined contrasts, then these and not the LSD should be used for mean separation. *True*
2. It is not important whether you include all of the interactions in your model, your *F*-test result for each effect will be identical in the model with and without the interactions. For example, if a study includes genotypes and compost rates, we can ignore the interaction of Genotype × Compost and each p-value for Genotype and Compost will be identical to each p-value for Genotype and Compost in the analysis that includes the Genotype × Compost interaction. *False*. By including interactions, the residual variance will change, therefore each term in the Analysis of Variance that included interactions will have a different F value compared with the analysis that did not include interactions.
3. In a study with three factors, the initial analysis must include main effects, two-way interactions, and the three-way interaction. If the three-way interaction is significant, then the researcher cannot properly discuss any main effect or two-way interaction. *False*. It is likely that the researcher will report results based on the analysis of the three-way interaction and we encourage researchers to follow this approach. However, if the significant *F* value for the three-way interaction is based largely on non-crossover interactions, then results for some main effects and some two-way interactions may provide similar information in a simpler format than the three-way interaction. When this is the case, the author may choose to present the data of the lower order effect as long as it is done so within the context of the higher-order interaction. When presented with this situation, there is not a standard correct approach. The author must determine which approach best (most accurately and in the simplest format) makes the points he/she feels are important.
4. It is never appropriate to speculate about what causes a significant effect. *False*.

Researchers should carefully identify results that lead to logical speculation based on their subject-area knowledge. It is crucial however, that the author clearly identifies these comments as speculation.

5. A researcher conducts an experiment in each of two years. When analyzing

such an experiment, it is always necessary that Year be considered as a random effect. *False*. Often when there are a low number of levels for what is often considered as a random effect, it is preferable to analyze it as a fixed effect.

6. The more levels a researcher has for an effect that is normally considered as a random effect (such as Year or Location), the more likely it would be useful to analyze it as a random effect. *True*

7. It is not important when making inferences whether an effect is fixed or random. *False*.

Inferences about a random effect should pertain to the population being tested. Inferences about fixed effects should pertain only to the specific effects (qualitative factor) or range of effects (quantitative factor) that were tested.

Exercise 1

Using the data corresponding to the example shown in Section 1 from this chapter, construct a SAS program that calculates in a step-wise fashion the F values for the main effects, 2-way, 3- way, and then the 4-way interactions as well as the single degree of freedom contrasts for the linear, quadratic, and cubic responses of GY to P for all effects involving P. Rather than using the method that we provided in Appendices 1 and 2 that automatically calculates the coefficients, in this exercise we ask that you insert the actual coefficients into each line of code.

Answer

The first step is to run Proc IML to obtain the correct coefficients. Since our unequally spaced rates of P are 0, 50, 150, and 250 kg P ha^{-1}, the following code will produce the correct regression coefficients.

```
Proc IML;
      Pcoeff=Orpol({0,50,150,250});
      Print Pcoeff;
Run;
```

The output from this program is the following:

Pcoeff			
0.5	-0.58585	0.4959593	-0.401004
0.5	-0.325472	-0.280609	0.7518821
0.5	0.1952834	-0.678681	-0.501255
0.5	0.716039	0.4633304	0.1503764

The numbers in the second column in this table are the coefficients needed to calculate a linear regression for the four rates of P we input. The third column provides the coefficients for a quadratic regression, and the final column contains the coefficients needed to calculate a cubic regression. In the code we provided in Appendices 1 and 2, the numbers in the second column (linear) are the "PosCoefGrade1", the numbers in the third column are the "PosCoefGrade2", and the numbers in the final column are the "PosCoefGrade3". To obtain the NegCoefGrade1, NegCoefGrade2, and NegCoefGrade3, we multiplied each column by (-1).

Now that we have our linear, quadratic, and cubic regression coefficients, our next step will be to read in our data and print it out to check that it has been read in correctly.

```
**** Reading data: Data Wheat is available in attached CSV file ****;
Data Wheat;
```

```
      Infile "C:\Experiment 1 Data Wheat.CSV" dlm="," firstobs=3 ;
      Informat Soil$ 10. ;
      Input Year Soil N P Rep Yield ; Datalines;
Run;
```

The next step is to insert the Proc GLIMMIX code for this data with four fixed-effect factors. In this first set of code, there will be only main effects. The order of the variables in the Class statement is particularly important. For our SAS code that generates the coefficients, P must follow all other fixed variables. If Year, Soil, or N had different numbers of levels, then their order would also be important. Since they all have the same number of levels (two), their order is not important except that they must be listed prior to P.

```
ODS Select CovParms Tests3 Contrasts; Proc GLIMMIX data = Wheat;
      class Year Soil N P Rep ;
      model Yield = Year Soil N P ;

      random Rep(Year Soil);
Run;
```

Our code for the three (linear, quadratic, and cubic) contrasts for the main effect of P follow. We use the coefficients from Columns 2, 3, and 4, respectively for each of these contrasts. The code for the linear effect of P is:

```
contrast "Linear P" P -0.58585 -0.325472 0.1952834 0.716039;
```

The words "Linear P" comprise a user-chosen title to describe the effect being tested. Following the closed quotation marks is P which tells SAS that all calculations will be done on this effect. The linear coefficients follow the P and the ";" ends the statement. Similarly, we calculate the quadratic and cubic effects as follows:

```
contrast "Quadratic P" P 0.4959593 -0.280609 -0.678681 0.4633304;
contrast "Cubic P" P -0.401004 0.7518821 -0.501255 0.1503764;
```

We insert a run statement after the final contrast statement and combine our Proc GLIMMIX code from above.

```
ODS Select CovParms Tests3 Contrasts;
Proc GLIMMIX data = Wheat;
      class Year Soil N P Rep ;
      model Yield = Year Soil N P ;
      random Rep(Year Soil);
      contrast "Linear P" P -0.58585 -0.325472 0.1952834 0.716039;
      contrast "Quadratic P" P 0.4959593 -0.280609 -0.678681 0.4633304;
      contrast "Cubic P" P -0.401004 0.7518821 -0.501255 0.1503764;
Run;
```

In addition to the standard Proc GLIMMIX output, running this code will result in the following output for these three Contrast statements.

Contrasts

Label	Num DF	Den DF	F Value	Pr > F
Linear P	1	52	25.94	<0.0001
Quadratic P	1	52	3.06	0.0860
Cubic P	1	52	0.44	0.5093

Now, we need to delete the last run statement and add our SAS code for the two-way interactions. Everything except the new contrast statements follows:

```
ODS Select CovParms Tests3 Contrasts; Proc GLIMMIX data = Wheat;
class Year Soil N P Rep ;
model Yield = Year Soil N P
      Year*Soil Year*N Year*P Soil*N Soil*P N*P ;
random Rep(Year Soil);
contrast "Linear P" P -0.58585 -0.325472 0.1952834 0.716039; contrast
```

```
"Quadratic P" P 0.4959593 -0.280609 -0.678681 0.4633304; contrast
"Cubic P" P -0.401004 0.7518821 -0.501255 0.1503764;
```

Now we need to construct and add the contrast statements for all 2-way interactions involving P.

```
contrast "Year × P Linear" Year*P
     -0.58585 -0.325472 0.1952834 0.716039
      0.58585 0.325472 -0.1952834 -0.716039;
```

Because we are analyzing the linear effect of P for the Year × P interaction, in the first line of code, after the title of our interaction within quotation marks, we insert the code Year*P to direct SAS to apply the coefficients to this interaction. The first line of coefficients are from the second column of the Proc IML output because this is a linear effect. These coefficients correspond to the code "PosCoefGrade1" in Appendices 1 and 2. The second line of coefficients are the coefficients in the first line multiplied by (-1) and these correspond to "NegCoefGrade1". Note that the ";" follows the second line of coefficients. We could have put these three lines of code without carriage returns as follows:

```
contrast "Year × P Linear" Year*P
     -0.58585 -0.325472 0.1952834 0.716039
       0.58585 0.325472 -0.1952834 -0.716039;
```

which is equivalent to the statement

```
*&PosCoefGrade1 &NegCoefGrade1; when using the SAS Macro code;
```

However, we recommend using the separate lines to facilitate finding errors in code.

If we add all of our contrast statements for 2-way interactions involving P, then we will have the following code.

```
ODS Select CovParms Tests3 Contrasts;
Proc GLIMMIX data = Wheat;
      class Year Soil N P Rep ;
      model Yield = Year Soil N P
           Year*Soil Year*N Year*P Soil*N Soil*P N*P;
      random Rep(Year Soil);
      contrast "Linear P" P -0.58585 -0.325472 0.1952834 0.716039;
      contrast "Quadratic P" P 0.4959593 -0.280609 -0.678681 0.4633304;
      contrast "Cubic P" P -0.401004 0.7518821 -0.501255 0.1503764;
      contrast "Year × P Linear" Year*P
          -0.58585 -0.325472 0.1952834 0.716039
           0.58585 0.325472 -0.1952834 -0.716039;
* First line is the original linear coefficients from Proc IML output;
*Second line is the Proc IML linear coefficients multiplied by -1;
*Note that P must come after Soil in the Class statement for this to
calculate properly;
*&PosCoefGrade1 &NegCoefGrade1;
contrast "Year × P Quadratic" Year*P
      0.4959593 -0.280609 -0.678681 0.4633304
     -0.4959593 0.280609 0.678681 -0.4633304;
*&PosCoefGrade2 &NegCoefGrade2 ;
contrast "Year × P Cubic" Year*P
     -0.401004 0.7518821 -0.501255 0.1503764
      0.401004 -0.7518821 0.501255 -0.1503764;
*&PosCoefGrade3 &NegCoefGrade3 ;
contrast "Soil × P Linear" Soil*P
     -0.58585 -0.325472 0.1952834 0.716039
      0.58585 0.325472 -0.1952834 -0.716039;
contrast " Soil × P Quadratic" Soil*P
      0.4959593 -0.280609 -0.678681 0.4633304
     -0.4959593 0.280609 0.678681 -0.4633304;
contrast "Soil × P Cubic" Soil*P
     -0.401004 0.7518821 -0.501255 0.1503764
      0.401004 -0.7518821 0.501255 -0.1503764;
contrast "N × P Linear" N*P
```

```
     -0.58585 -0.325472 0.1952834 0.716039
      0.58585 0.325472 -0.1952834 -0.716039;
contrast "N × P Quadratic" N*P
      0.4959593 -0.280609 -0.678681 0.4633304
     -0.4959593 0.280609 0.678681 -0.4633304;
contrast "N × P Cubic" N*P
     -0.401004 0.7518821 -0.501255 0.1503764
      0.401004 -0.7518821 0.501255 -0.1503764;
```

As noted above, and for all other interactions, note that in the Class statement, P must have followed whatever fixed effect(s) that comprise the interaction, in this case Year, Soil, and N. This is because we calculated these coefficients as follows:

For P, we used the coefficients generated by Proc IML, and for Year we used +1, -1, for Year 1 and Year 2, respectively. Note that when you are using rates, such as 0, 50, 150, and 250, SAS will assign the coefficients to these rates from lowest to highest rate. For Year 1 or Year 2, SAS will assign the coefficients 1 -1 alphabetically so Year 1 will correspond to 1 and -1 to Year 2. Had the treatment been month, and the two months in question January and August, then 1 would have been assigned to August and -1 would have been assigned to January. The output follows for all of the main effects and two-way interaction contrasts that include P as one of the factors:

Contrasts

Label	Num DF	Den DF	F Value	Pr > F
Linear P	1	41	40.42	< 0.0001
Quadratic P	1	41	4.77	0.0347
Cubic P	1	41	0.69	0.4117
Year × P Linear	1	41	5.76	0.0210
Year × P Quadratic	1	41	1.93	0.1725
Year × P Cubic	1	41	0.09	0.7648
Soil × P Linear	1	41	16.60	0.0002
Soil × P Quadratic	1	41	0.07	0.7880
Soil × P Cubic	1	41	0.00	0.9877
N × P Linear	1	41	0.98	0.3280
N × P Quadratic	1	41	0.18	0.6710
N × P Cubic	1	41	0.52	0.4729

Note that now by adding the two-way interactions, the linear, quadratic, and cubic responses to the main effect of P changed because the denominator degrees of freedom decreased from 52 to 41.

Based on the order of variables in the Class statement, we calculated the 8 coefficients for the Year × P interaction using the following table.

Phosphorus fertilizer rates

Year treatment	Year coefficient	kg P ha^{-1}				
		0	50	150	250	Sum of coefficients
		Coefficients for linear response to P				
		-0.58585	-0.325472	0.1952834	0.716039	0.0000004
		Coefficients for Year × P linear interaction				
Year 1	1	-0.58585	-0.325472	0.1952834	0.716039	0.0000004
Year 2	-1	0.58585	0.325472	-0.1952834	-0.716039	-0.0000004
Sum of coefficients	0	0.00000	0.000000	0.00000000	0.00000	0.0000000

The table above shows how we multiplied the P coefficients by the Year coefficients to obtain the coefficients used in the SAS code for the linear response to P in the Year × P interaction.

Note that the sum of the coefficients in the rows for Year (Year 1 and Year 2) do not sum exactly to 0. However, they are extremely close to 0. In order to ensure that the contrasts are calculated correctly, it is important not to round off the coefficients generated by Proc IML.

If the Class statement would have had P listed before Year, then we would have calculated our coefficients using the following table:

P fertilizer		Year coefficients	
kg ha^{-1}	Coefficients for linear response to P	Year 1 1	Year 2 -1
0	-0.5858500	-0.5858500	0.5858500
50	-0.3254720	-0.3254720	0.3254720
150	0.1952834	0.1952834	-0.1952834
250	0.7160390	0.7160390	-0.7160390
Sum of coefficients	0.0000004	0.0000004	0.0000004

From this table, we could have run the following program to calculate the linear response to P fertilizer for the Year × P Interaction: Notice that P is listed before Year in the Class statement below.

```
Proc GLIMMIX data = Wheat;
     class Soil N P Year Rep ;
     model Yield = Year Soil N P
          Year*Soil Year*N Year*P Soil*N Soil*P N*P ;
     random Rep(Year Soil);
     contrast " Year × P Linear" Year*P
         -0.5858500 0.5858500
         -0.3254720 0.3254720
          0.1952834 -0.1952834
          0.7160300 -0.7160300
Run;
```

Our output for the contrast Year × P Linear follows:

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
Year × P Linear	1	45	5.76	0.0206

The output is the same as previously when P came after instead of before Year in the Class statement. The values can be calculated correctly either way. (Verify this on your own.) However, the researcher needs to be aware of the order of variables in the Class statement so he/she can properly order the coefficients in the SAS code.

As long as the coefficients are listed in an order that is in agreement with the variables in the Class statement, then the contrast will be calculated correctly.

Next, we calculate the coefficients for the three-way single degree of freedom interactions that include P, beginning with the Year × Soil × P interaction.

Beginning with Year and Soil, we assign the following coefficients:

Year 1 = +1

Year 2 = -1

Black Soil = +1

Chestnut Soil = -1

We have a simple table with +1 and -1 being the coefficients on the row and +1 and -1 being the coefficients on the column. Multiplying the Year × Soil coefficients we get the following 4 coefficients: 1, -1, -1, 1 shown in yellow in the following table.

Year	Year		Soil	
	Coefficients	Black	Chestnut	
		Soil coefficients		Sum of coefficients
		1	-1	0
Year 1	1	1	-1	0
Year 2	-1	-1	1	0
Sum of coefficients	0	0	0	0

The numbers shaded in yellow are the products of the Year × Soil coefficients, and the sums in the bottom row are of the two rows immediately above the bottom, and in the final column, the sums are of the two columns immediately to the left of the final column.

The coefficients 1, -1, -1, and 1 are then placed in the rows with the P linear coefficients forming the columns (we placed P after Year and Soil in the Class statement) to calculate the contrast for the linear response to P in the Year × Soil x P interaction.

Year × Soil	Coefficients for linear response of P linear				
Coefficients	-0.58585	-0.325472	0.1952834	0.71603	Sum
1	-0.58585	-0.325472	0.1952834	0.716039	-0.0000004
-1	0.58585	0.325472	-0.1952834	-0.716039	0.0000004
-1	0.58585	0.325472	-0.1952834	-0.716039	0.0000004
1	-0.58585	-0.325472	0.1952834	0.716039	-0.0000004
Sum	0.00000	0.000000	0.0000000	0.00000	0.0000000

The numbers shaded in yellow are the products of the coefficients for the linear response to P in the Year × Soil × P interaction. These are the numbers that we will use as coefficients in the SAS code for the Contrast statement that calculates the linear response to P in each of our three-way interactions. The coefficients for the quadratic and cubic responses to P are calculated similarly. The following code calculates linear, quadratic, and cubic responses to P in each three-way interaction involving P as a factor.

```
Proc GLIMMIX data=Wheat;
      class Year Soil N P Rep;
      model Yield=Year Soil N P
      Year*Soil Year*N Year*P Soil*N Soil*P N*P
      Year*Soil*N Year*Soil*P Year*N*P Soil*N*P;
random Rep(Year Soil);
**** Year × Soil × P three way interaction contrasts ****;
contrast "Year × Soil × P Linear" Year*Soil*P
      0.58585 0.325472 -0.1952834 -0.716039
     -0.58585 -0.325472 0.1952834 0.716039
     -0.58585 -0.325472 0.1952834 0.716039
      0.58585 0.325472 -0.1952834 -0.716039;
contrast "Year × Soil × P Quadratic" Year*Soil*P
     -0.4959593 0.280609 0.678681 -0.4633304
      0.4959593 -0.280609 -0.678681 0.4633304
      0.4959593 -0.280609 -0.678681 0.4633304
     -0.4959593 0.280609 0.678681 -0.4633304;
contrast "Year × Soil × P Cubic" Year*Soil*P
      0.401004 -0.7518821 0.501255 -0.1503764
     -0.401004 0.7518821 -0.501255 0.1503764
     -0.401004 0.7518821 -0.501255 0.1503764
      0.401004 -0.7518821 0.501255 -0.1503764;
contrast "Year × N × P Linear" Year*N*P
      0.58585 0.325472 -0.1952834 -0.716039
     -0.58585 -0.325472 0.1952834 0.716039
     -0.58585 -0.325472 0.1952834 0.716039
      0.58585 0.325472 -0.1952834 -0.716039;
contrast "Year× N × P Quadratic" Year*N*P
     -0.4959593 0.280609 0.678681 -0.4633304
      0.4959593 -0.280609 -0.678681 0.4633304
      0.4959593 -0.280609 -0.678681 0.4633304
     -0.4959593 0.280609 0.678681 -0.4633304;
contrast "Year × N × P Cubic" Year*N*P
      0.401004 -0.7518821 0.501255 -0.1503764
```

```
     -0.401004 0.7518821 -0.501255 0.1503764
     -0.401004 0.7518821 -0.501255 0.1503764
      0.401004 -0.7518821 0.501255 -0.1503764;
contrast "Soil × N × P Linear" Soil*N*P
      0.58585 0.325472 -0.1952834 -0.716039
     -0.58585 -0.325472 0.1952834 0.716039
     -0.58585 -0.325472 0.1952834 0.716039
      0.58585 0.325472 -0.1952834 -0.716039;
contrast "Soil × N × P Quadratic" Soil*N*P
     -0.4959593 0.280609 0.678681 -0.4633304
      0.4959593 -0.280609 -0.678681 0.4633304
      0.4959593 -0.280609 -0.678681 0.4633304
     -0.4959593 0.280609 0.678681 -0.4633304;
contrast "Soil × N × P Cubic" Soil*N*P
      0.401004 -0.7518821 0.501255 -0.1503764
     -0.401004 0.7518821 -0.501255 0.1503764
     -0.401004 0.7518821 -0.501255 0.1503764
      0.401004 -0.7518821 0.501255 -0.1503764;
Run;
```

The results for the contrasts of each three-way interaction follow.

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
Year × Soil × P Linear	1	31	22.49	< 0.0001
Year × Soil × P Quadratic	1	31	0.02	0.8847
Year × Soil × P Cubic	1	31	0.45	0.5053
Year × N × P Linear	1	31	0.74	0.3955
Year × N × P Quadratic	1	31	0.44	0.5118
Year × N × P Cubic	1	31	0.58	0.4536
Soil × N × P Linear	1	31	2.70	0.1103
Soil × N × P Quadratic	1	31	0.07	0.7970
Soil × N × P Cubic	1	31	11.79	0.0017

To finish the exercise, it is necessary to calculate the three single degree of freedom contrasts for the four way interaction Year × Soil × N × P. For this set of contrasts, we will provide the detail for calculating the coefficients for the cubic response of P for this four-way interaction. From our calculations for the coefficients used in the three-way interactions, we know that the coefficients for the Year × Soil interaction are 1, -1, -1, and 1. To obtain the coefficients for the three-way Year × Soil × N interaction, we place these four coefficients as rows and for N, add 1 and -1 as columns, giving us the following table.

Year × Soil coefficients		N coefficients	Sum of coefficients
	1	-1	
1	1	-1	0
-1	-1	1	0
-1	-1	1	0
1	1	-1	0
Sum of coefficients	0	0	0

By adding the factor N which had two levels, we have evolved from four coefficients for the Year × Soil two-way interaction to eight coefficients for the Year × Soil × N three-way interaction. The coefficients which are the product of the rows and columns in the above table are 1, -1, -1, 1, -1, 1, 1, and -1. To obtain our coefficients for the four-way interaction, we build a new table using these 8 coefficients as the rows and (since we are calculating the cubic response to P rates) using the cubic coefficients for our rates of P generated previously by Proc IML as the columns. The results are in the following table.

Y × S × N coefficients	Cubic coefficients for P from Proc IML				Sum of coefficients
	-0.401004	0.7518821	-0.501255	0.1503764	-0.0000005
1	-0.401004	0.7518821	-0.501255	0.1503764	-0.0000005
-1	0.401004	-0.7518821	0.501255	-0.1503764	0.0000005
-1	0.401004	-0.7518821	0.501255	-0.1503764	0.0000005
1	-0.401004	0.7518821	-0.501255	0.1503764	-0.0000005
-1	0.401004	-0.7518821	0.501255	-0.1503764	0.0000005
1	-0.401004	0.7518821	-0.501255	0.1503764	-0.0000005
1	-0.401004	0.7518821	-0.501255	0.1503764	-0.0000005
-1	0.401004	-0.7518821	0.501255	-0.1503764	0.0000005

The coefficients necessary to calculate the significance of the Year × Soil × N × P Cubic interaction are those in yellow:

Getting back to the code we provided in Appendix 2, to calculate the significance of this interaction, we have the following groups:

```
-0.401004   0.7518821   -0.501255    0.1503764:   "PosCoefGrade3"
 0.401004   -0.7518821   0.501255   -0.1503764:   "NegCoefGrade3"
 0.401004   -0.7518821   0.501255   -0.1503764:   "NegCoefGrade3"
-0.401004   0.7518821   -0.501255    0.1503764:   "PosCoefGrade3"
 0.401004   -0.7518821   0.501255   -0.1503764:   "NegCoefGrade3"
-0.401004   0.7518821   -0.501255    0.1503764:   "PosCoefGrade3"
-0.401004   0.7518821   -0.501255    0.1503764:   "PosCoefGrade3"
 0.401004   -0.7518821   0.501255   -0.1503764:   "NegCoefGrade3"
```

We would calculate the linear (PosCoefGrad1" and "NegCoefGrade1") and quadratic coefficients (PosCoefGrad2" and "NegCoefGrade2") similarly, but using the second and third columns, respectively, from the Proc IML output instead of the fourth column which we used to generate these cubic coefficients.

The code for calculating the linear, quadratic, and cubic responses to P for the four-way interaction Year × Soil × N × P is the following:

```
Proc GLIMMIX data = Wheat;
class Year Soil N P Rep ;
model Yield = Year Soil N P
              Year*Soil Year*N Year*P Soil*N Soil*P N*P
              Year*Soil*N Year*Soil*P Year*N*P Soil*N*P
              Year*Soil*N*P;
random Rep(Year Soil);
contrast "Year × Soil × N × P Linear" Year*Soil*N*P
     -0.58585 -0.325472 0.1952834 0.716039
      0.58585 0.325472 -0.1952834 -0.716039
      0.58585 0.325472 -0.1952834 -0.716039
     -0.58585 -0.325472 0.1952834 0.716039
      0.58585 0.325472 -0.1952834 -0.716039
     -0.58585 -0.325472 0.1952834 0.716039
     -0.58585 -0.325472 0.1952834 0.716039
      0.58585 0.325472 -0.1952834 -0.716039;
contrast "Year × Soil × N × P Quadratic" Year*Soil*N*P
      0.4959593 -0.280609 -0.678681 0.4633304
     -0.4959593 0.280609 0.678681 -0.4633304
     -0.4959593 0.280609 0.678681 -0.4633304
      0.4959593 -0.280609 -0.678681 0.4633304
     -0.4959593 0.280609 0.678681 -0.4633304
      0.4959593 -0.280609 -0.678681 0.4633304
      0.4959593 -0.280609 -0.678681 0.4633304
     -0.4959593 0.280609 0.678681 -0.4633304;
contrast "Year × Soil × N × P Cubic" Year*Soil*N*P
     -0.401004 0.7518821 -0.501255 0.1503764
      0.401004 -0.7518821 0.501255 -0.1503764
      0.401004 -0.7518821 0.501255 -0.1503764
```

```
     -0.401004 0.7518821 -0.501255 0.1503764
      0.401004 -0.7518821 0.501255 -0.1503764
     -0.401004 0.7518821 -0.501255 0.1503764
     -0.401004 0.7518821 -0.501255 0.1503764
      0.401004 -0.7518821 0.501255 -0.1503764;
Run;
```

The output for the linear, quadratic, and cubic responses to P for the four-way interaction follows:

Contrasts				
Label	Num DF	Den DF	F Value	Pr > F
Year × Soil × N × P Linear	1	28	0.50	0.4870
Year × Soil × N × P Quadratic	1	28	0.01	0.9436
Year × Soil × N × P Cubic	1	28	5.90	0.0218

The complete Proc GLIMMIX SAS code for calculating all two-way, three-way, and four-way interactions in the ANOVA as well as the linear, quadratic, and cubic responses to P for all contrasts involving P, can be easily obtained putting together the programs shown above. Note: *using Proc GLIMMIX, instead of the CONTRAST statement you can also use the LSMESTIMATE statement because it provides a mechanism for obtaining hypothesis tests among the least squares means. However, in contrast to the hypotheses tested with the ESTIMATE or CONTRAST statements, the LSMESTIMATE statement enables to form linear combinations of the least squares means, rather than linear combination of fixed-effects parameter estimates and/or random-effects solutions*

Exercise 2

Using the data set (Wheat) shown in Section 1 from this chapter, calculate the Mean Squared Error (MSE) for Rep (Year Soil) for the complete model using two approaches.

Approach 1.

Proc GLIMMIX provides the covariance parameter estimate (or variance) for Rep (Year Soil). Referring to Expected Mean Squares, we know the following:

```
MSE for Rep (Year Soil) = Residual variance + (P × N) × [Variance
(Rep (Year × Soil)]
Where P and N equal the number of levels of P and N, respectively.
P × N = 4 × 2 = 8
MSE for Rep (Year Soil) = Residual variance + (8) × [Variance (Rep
(Year × Soil)]
```

We know from the Proc GLIMMIX output that Variance (Rep (Year × Soil) = 0.000292 and we know that the residual variance is 0.008036. Thus,

```
MSE for Rep (Year × Soil) = 0.008036 + 8(0.000292)
MSE for Rep (Year × Soil) = 0.008036 + 0.002336
MSE for Rep (Year × Soil) = 0.010372
```

Approach 2.

We can also use our LSD output in Appendix 2, Example 3, from Proc GLIMMIX to calculate the MSE for Rep (Year Soil). One column of output is the average standard error of the differences between two means (ASED) which is labeled "AvStdErr" in the Proc GLIMMIX output. We know that the MSE for Rep (Year Soil) is used to calculate the standard error of the difference (SED) between two Year, Soil, or Year × Soil means, so we can use that formula in reverse.

Using the output for Year × Soil (we could also have used Year or Soil because, as stated above, the ASED for each of these was calculated from the same Mean Square as was used to calculate the LSD for Year × Soil).

The actual formula for the SED is SED = [2(MSE for Rep (Year Soil))/n](1/2) where n = number of observations for each Year × Soil mean. This number n is calculated from 2 reps, 4 rates of P, and 2 rates of N so n = 2 × 4 × 2 = 16 observations for calculating each Year × Soil mean. Substituting ASED (which is AvStdErr from the Proc GLIMMIX output) for SED, this gives us:

```
0.036011 = [2(MSE for Rep (Year Soil))/16]^(1/2)
```

Squaring both sides of the equation gives us:

```
0.0012967 = 2(MSE for Rep (Year Soil)/16
0.0012967 = MSE for Rep (Year Soil)/8
8 × 0.0012967 = MSE for Rep (Year Soil)
```

MSE for Rep (Year Soil) = 0.0103736, which is equal to 5 decimal places to the answer from our Approach 1.

Calculating LSD values

Now, using the residual variance (MSE) or the mean square of Rep (Year Soil) if appropriate, let's calculate the LSD for all effects and interactions and compare these LSDs with the LSDs calculated with Proc GLIMMIX, which are provided in the output in Appendix 2. Remember, the LSD is a mechanism for doing all possible t tests with a mean variance. However, the Pdiff option in SAS actually calculates all possible t tests and does so not based on a mean variance, but based on the actual pooled variance of the difference of the two means being tested. Thus, using the Pdiff option is a more precise way of carrying out the intentions of the LSD so we recommend, if the results differ, using the significance results from the Pdiff option rather than from the LSD. Still, when we publish, it is useful to provide measures of variability in our tables and figures. The LSD is a useful statistic to use for this purpose. While the actual probabilities may not be exactly equal, comparisons among every single pair of means for significance using the LSD will almost always yield the same significance results (that is, significant or not significant) as the Pdiff option. For rare instances when the Pdiff option identifies a comparison as significant or not significant and this determination is different from that of the LSD, the author can point this out and indicate that it is the result from the Pdiff options on which conclusions are based, but that the LSD value is presented because it is still a useful estimate of variation.

For our calculations, we need to remember that the general formula for calculating the LSD is for t (α, df) (and we will conduct our calculations at $\alpha = 0.05$),

```
LSD = t [(2 × variance)/n]^(1/2)
```

Where n = the number of observations for each mean.

We will show the calculations for calculating the LSDs for comparing means of N, Soil, Year × Soil, and Year × N and results for all effects and interactions are provided in the table below. We need to recall that each effect had the following number of levels.

```
Year = 2
Soil = 2
N = 2
P = 4
Rep = 2
```

First for N.

The two N means were comprised of 2 × 2 × 4 × 2 = 32 observations (number of levels of Year, Soil, P, and Rep multiplied together), so n = 32.

Based on our Proc GLIMMIX output from Appendix 2, we see that the denominator df associated with N = 28. This tells us two things. First, the appropriate variance is the residual (error mean square) since there are 28 df associated with this value. Second, to find the appropriate t value, we need to use df = 28.

We see from the SAS output that the mean square error = 0.008036 and from a t table, we find that t = 2.048. (We can also use the following function in Excel to obtain our *t* value:

```
=tinv (0.05, 28) = 2.048407
```

We now have the information to calculate the LSD which is:

LSD = 2.048407 [2 x 0.008036)/32]$^{(1/2)}$

LSD = 0.045906. This compares with 0.04591 calculated by using the ASED estimated by Proc GLIMMIX.

LSD for comparing means of Soil.

This is equivalent to the LSD for comparing means of Year because there are two levels of Year and two levels of Soil, and to calculate the LSD for year or soil, the variance is the same and it is the MSE for Rep (Year Soil) = 0.010372 as calculated above.

As for N, there are 32 observations that make up each Soil mean (Year × N × P × Rep) = 2 × 2 × 4 × 2 = 32.

Unlike the example of calculating an LSD for comparing N means, where denominator df = 28, we see from our Proc GLIMMIX output that for Soil (or Year), the denominator df = 4. Thus, we must use the mean square for Rep (Year Soil), which we already calculated as 0.010372, in our LSD formula and we must use 4 df for finding our t value which is *t* = 2.776445. Thus,

```
LSD = 2.776445 [2 × 0.010372)/32] (1/2)
LSD = 2.776445 × 0.0254597
LSD = 0.0706903.
```

This compares with 0.07069 calculated in Appendix 2 by using the ASED estimated by Proc GLIMMIX.

LSD for comparing means of the Year × Soil interaction.

There are 16 observations that comprise each Year × Soil mean (Levels of N × P × Rep) = 2 × 4 × 2 = 16. We see that the denominator df = 4 for Year × Soil, thus we will use the mean square error of Rep (Year Soil) as the variance, which is 0.010372 and from the t table, we find that t = 2.77645. Thus,

```
LSD = 2.776445 × [2 × 0.010372/16] (1/2)
LSD = 2.776445 × 0.0360069
LSD = 0.09997129
```

which compares with 0.09997 calculated by using the ASED estimated by Proc GLIMMIX.

LSD for comparing means of the Year × N interaction.

There are 16 observations that comprise each Year × N mean (Levels of Soil × P × Rep) = 2 × 4 × 2 = 16. We see that the denominator df = 28 for Year × N, thus we will use the residual as the variance = 0.008036, and from the t table, we find that t = 2.048407. Thus,

```
LSD = 2.048407 [2 × 0.008036/16] (1/2)
LSD = 2.048407 × 0.0316938
LSD = 0.0649219
```

which compares with 0.06796 calculated in Appendix 2, Example 3 by using the ASED estimated by Proc GLIMMIX.

The table below shows that our LSD values were nearly identical to 5 decimal places, whether calculated in Appendix 2 by the ASED or here using the variance. In most cases, it should be fine to use the correct estimated variance in order to estimate the LSD. To be most certain of precision, it would be safest to use Proc GLIMMIX and the ASED method we provided in Appendix 2.

Effect	Variance	ASED	Den df†	Critical t‡	LSD by	LSD by
Year (Y)	0.010372	0.02546	4	2.77645	0.07069	0.07069
Soil (S)	0.010372	0.02546	4	2.77645	0.07069	0.07069
Nitrogen (N)	0.008036	0.02241	28	2.04841	0.04590	0.04591
Phosphorus (P)	0.008036	0.02546	28	2.04841	0.06492	0.06492
Y × S	0.010372	0.03601	4	2.77645	0.09997	0.09997
Y × N	0.008036	0.03318	28	2.04841	0.06492	0.06796
Y × P	0.008036	0.04574	28	2.04841	0.09181	0.09369
S × N	0.008036	0.03318	28	2.04841	0.06422	0.06796
S × P	0.008036	0.04574	28	2.04841	0.09181	0.09369
N × P	0.008036	0.04482	28	2.04841	0.09181	0.09181
Y × S × N	0.008036	0.04752	28	2.04841	0.09181	0.09734
Y × S × P	0.008036	0.13355	28	2.04841	0.12984	0.13355
Y × N × P	0.008036	0.06400	28	2.04841	0.12984	0.13109
S × N × P	0.008036	0.06400	28	2.04841	0.12984	0.13109
Y × S × N × P	0.008036	0.09089	28	2.04841	0.18363	0.18629

† Denominator degrees of freedom from the ANOVA for the fixed effects
‡ Critical t value at 0.05/2 significance level and for Den df.

CHAPTER 8: THE ANALYSIS OF COMBINED EXPERIMENTS

Philip M. Dixon, Kenneth J. Moore, and Edzard van Santen

Answers to Review Questions, with short explanations

1. In this study, modeling environment effects and block effects as fixed effects or as random effects leads to the same inference about the difference between the two types of tillage.

   True. The design is balanced. The only differences would arise if a random effect variance is estimated as 0.

2. If conclusions about the effect of tillage in a new location are desired, you should use narrow sense inference.

   False. Narrow sense inference makes conclusions about the locations and years used in the study

3. Narrow sense conclusions about the effects of tillage usually have smaller standard errors than do broad sense conclusions.

   True. The variance for broad sense conclusions includes an additional non-negative variance component.

4. To obtain narrow-sense conclusions, omit the treatment by environment interaction from the model.

   False. You should model the treatment by environment interaction as a fixed effect. Omitting it from the model assumes no interaction and pools the interaction with the plot-plot variation.

5. To obtain broad-sense conclusions, model the treatment by environment interaction as a random effect.

   True.

6. Broad-sense confidence intervals for the difference between two types of tillage will be based on T distributions with 18 degrees of freedom.

   False. The degrees of freedom for the confidence interval will be the df of the treatment by environment interaction, which is (18 – 1)(3-1) = 34

7. The combined analysis across environments requires that plot-plot variation be pooled across environments.

   False. Although error variances are often pooled, they do not need to be.

8. The combined analysis across environments requires that variation between blocks be pooled across environments.

   False. Although block variances are often pooled, they do not need to be. If blocks are considered a fixed effect, pooling is not an issue.

9. The 18 environments are actually 6 locations, each studied for 3 years. Tillage effects are expected to vary somewhat among locations because of different soil

characteristics. Tillage effects are not expected to vary among years at the same location. In this case, subdividing the treatment by environment interaction will have minimal effect on the conclusions about the average treatment effect.

True. Although the location by treatment variance component is expected to be somewhat different from the year by treatment variance component, this appears to have minimal effect on conclusions about the average treatment effect.

10. Imagine that the three tillage treatments are three levels of some quantitative factor, e.g. amount of soil disturbance. The data for each environment could be analyzed using a regression model with a linear effect of soil disturbance. It is possible to construct a combined analysis of those regression models in all 18 environments.

True. The model for the combined analysis includes a random regression slope by environment interaction.

Chapter 8 Answers to Exercises

1. An experiment was conducted to assess the effect of a fungicide treatment on soybean yield (kg ha^{-1}). It was conducted as an on-farm strip-plot trial with six pairs of side-by-side strips of which one randomly received fungicide treatment. The experiment was repeated at eleven farms (environments). The data were extracted from a much larger dataset provided by the Iowa Soybean Association and are provided in the on-farm soybean dataset in the supplemental materials.

a. Analyze the experiment separately for each environment.

See SAS code – first block of code

b. Evaluate the error variances to determine whether or not they may be considered homogeneous.

See SAS code – second block of code using repeated statement and reml to compare variances across environments

Various approaches are possible, including Levene's Test, a Likelihood Ratio Test, and comparison of AIC, AICc, or BIC statistics. All are consistent and indicate that the model with separate variances for each location is superior to using a single pooled estimate. The inference is that variances are heterogeneous.

c. Conduct a combined analysis of variance assuming environment and replication to be random factors and treatment as fixed.

See SAS code – third block of code

The F ratio for treatment is 13.78. Since the probability associated with its occurrence is quite small (0.0038) the effect of fungicide treatment is considered to be significant.

d. Interpret the results of the experiment with respect to the efficacy of fungicide treatment in improving soybean yield.

The mean difference between fungicide and untreated soybean over all environments was approximately 1.65 kg ha^{-1} (se = 0.44 kg ha^{-1}). To be economical on average for represented environments, the per-ha cost of application would need to be less than the market value of 1.65 kg of soybeans.

2. Antonio Mallorino at Iowa State University has studied corn response to P fertilization since 2002. The Prate.csv file contains 13 years of data from the SouthWest research farm.

The design is a RCBD with three blocks of five plots each. Four P rates (0, 28, 56, and 112 lb ac^{-1}) were used; the 0 level was replicated twice in each block. Blocks and plots can be considered independent across years. The response variable is yield in bu ac^{-1}.

Consider years to be a fixed factor and P rate (Prate) to be a continuous variable.

a. What sort of polynomial model is appropriate to describe yield response to Prate? Linear? quadratic? With one coefficient for all years or coefficients that differ among years?

See SAS code - first block of code. A reasonable model has a quadratic response to P fertilization, with different linear coefficients but a single quadratic term. The quadratic × year interaction is not significant, but when that term is omitted, the quadratic effect, the linear effect, and the linear × year interaction are all significant. There is no evidence of lack of fit in either the interaction or main effect terms.

Now consider years to be a random factor. Fit a quadratic model that allows the intercept and linear Prate coefficient to vary between years (but the quadratic coefficient is constant).

b. What is the equation that predicts yield as a function of Prate for a year not in the data set, e.g, 2015?

See SAS code - second block of code, Solution for Fixed Effects in the output.

Yield = 170.2 + 0.36 × Prate - 0.00214 × $Prate^2$.

c. What is the year-to-year variability in the linear Prate slope? Use the standard deviation to describe that variability.

See SAS code - second block of code. The slope variance is the UN(2,2) parameter in the SAS output. We want its square root, = 0.12.

d. Examine the residuals. Is it appropriate to use yield as the response variable, or should yield be transformed?

See SAS code - third block of code. A plot of residuals against predicted values indicates no need to transform the response variable. The plot shows no evidence of lack of fit and no change in residual variability with increases in predicted values.

e. Apply Levene's test to the residuals to assess whether the error variance differs among years.

See SAS code - fourth block of code. There is evidence of different error variances in different years. The p-value is 0.022.

f. Refit the model used in parts b-e with year-specific error variances. Do the answers to questions b and c change much?

See SAS code - fifth block of code. The answers are slightly different, but the differences are small. When fit with year-specific error variances, the equation is:

Yield = 170.3 + 0.35 × Prate - 0.00202 × Prate2

The variability in the slope between years is 0.11.

CHAPTER 9: ANALYSIS OF COVARIANCE

Kevin S. McCarter

Sample SAS code for examples and exercises

```
dm "log;clear";
options nodate pageno=1;

ods html close;
ods html;

*Create the SAS datasets for the examples and exercises;

data example_1;
input grp $ y x @@;
datalines;
A 71.3 12.7 A 63.4 13.3 A 55.0  8.6 A 54.0  7.3 A 54.6  8.2
A 47.7  6.4 A 49.1  7.0 A 88.1 14.2 A 59.4  8.6 A 70.5 10.7
B 63.0  7.6 B 80.9 13.4 B 78.7 10.3 B 85.1 14.8 B 78.5 13.6
B 73.0 13.5 B 53.0  5.3 B 76.3  9.9 B 68.7  9.4 B 84.9 14.2
;
run;
data example_2;
input grp $ y x @@;
datalines;
A 69.2 10.8 A  59.4 10.7 A 70.2 13.1 A 52.3  6.6 A 61.0 9.6
A 73.9 13.4 A  57.1  7.3 A 64.9 10.0 A 68.2 13.8 A 75.1 14.8
B 89.9 17.4 B 101.3 21.2 B 73.2 13.4 B 96.4 20.1 B 86.4 17.6
B 74.8 14.8 B  81.2 17.2 B 97.3 20.2 B 99.4 21.9 B 79.3 13.9
;
run;

data example_3;
input grp $ y x @@;
datalines;
A 81.2 14.1 A 58.7  8.8 A 47.4  5.7 A 49.4  5.5 A 66.1  9.1
A 72.5 14.6 A 71.1 12.7 A 53.5  6.2 A 62.2  8.0 A 68.5 12.4
B 56.6 15.9 B 57.5 17.9 B 75.6 21.6 B 68.5 19.9 B 58.0 14.1
B 57.7 15.1 B 62.6 16.8 B 73.9 20.6 B 77.0 20.9 B 44.4 13.0
;
run;

data example_4;
input grp $ y x @@;
datalines;
A 55.1  8.2 A 67.1 11.0 A 73.6 14.8 A 64.6  9.3 A 76.4 13.8
A 45.5  5.3 A 47.3  6.4 A 57.4  6.6 A 78.5 12.9 A 61.9  8.4
B 61.7  9.5 B 56.3 11.3 B 54.2 12.0 B 68.0  9.8 B 58.5 12.6
B 59.2 10.5 B 60.8 10.7 B 68.2  9.8 B 52.7 13.1 B 78.5  6.4
;
run;
data exercise_1;
input grp $ y x @@;
datalines;
A   49.4  5.2 A  72.2 13.2 A 61.6  9.2 A  63.4 11.3 A  71.0 12.3
A   49.6  7.2 A  56.6  8.3 A 61.1  6.7 A  66.1  9.7 A  71.9 12.0
B   95.3 20.2 B  96.7 20.5 B 75.8 12.5 B 102.6 21.9 B  78.5 15.3
B 101.4 20.8 B 100.8 21.3 B 88.2 17.2 B  79.1 14.4 B 108.0 21.2
;
run;
```

```
data exercise_2;
input grp $ y x @@;
datalines;
A 56.7  7.0 A 46.3  5.7 A 71.2 14.6 A 69.0 11.9 A 59.7 10.8
A 59.8 10.2 A 60.1  8.1 A 66.8 10.8 A 72.5 14.5 A 57.6  7.9
B 70.3 10.5 B 78.0 12.8 B 63.1  9.4 B 57.1  6.6 B 68.7 11.3
B 70.4  9.5 B 73.3  9.4 B 61.7  6.8 B 73.7 11.9 B 77.4 11.6
;
run;

data exercise_3;
input grp $ y x @@;
datalines;
A 54.8  8.9 A 54.6  8.3 A 52.5 10.0 A 68.2 12.6 A 59.6  9.4
A 76.8 13.2 A 56.3  7.9 A 71.2 15.0 A 62.0  9.2 A 68.0 13.8
B 54.4 14.7 B 55.8 19.1 B 69.4 20.3 B 51.7 15.3 B 67.2 21.3
B 43.9 12.2 B 71.6 21.6 B 66.7 20.5 B 42.4 13.4 B 55.7 15.9
;
data exercise_4;
input grp $ y x @@;
datalines;
A 73.9 14.4 A 46.7  6.1 A 48.9  6.3 A 83.0 14.6 A 80.7 15.0
A 73.1 12.1 A 54.8  7.2 A 53.4  7.3 A 79.3 13.5 A 60.6 11.5
B 68.6  9.6 B 66.9  9.9 B 53.5  5.0 B 79.6 13.3 B 82.0 13.2
B 72.4 12.2 B 67.6  9.9 B 77.4 13.4 B 70.6 10.0 B 75.4 11.1
;
run;

data exercise_5;
input grp $ y x @@;
datalines;
A 70.9 12.6 A 71.8 13.5 A 61.0 10.9 A 71.4 13.9 A 56.9  6.4
A 55.8  9.1 A 62.1  7.4 A 80.8 13.2 A 54.8  8.8 A 50.8  5.6
B 52.4 12.3 B 57.9 10.7 B 68.9 11.1 B 62.7 11.6 B 62.4 10.2
B 78.5  6.5 B 55.0 14.8 B 79.2  6.8 B 82.6  5.5 B 51.6 13.7
;
run;
* Create SAS macro for performing analyses;

%macro analyze_data(DATASET);

footnote1 "Chapter 9 - Analysis of Covariance - Sample SAS Code for
Examples and Exercises";
footnote2 "Analysis of Dataset &DATASET";

* Print listing of dataset;

title1 "Listing of the Dataset &DATASET";
proc print data=&DATASET;
run;

* Calculate summary statistics and perform exploratory analysis;

title1 "Summary of the Dataset";
proc means data=&DATASET mean std;
var y x;
by grp;
run;

title1 "Boxplots of Response Variable Y for Each Group";
proc sgplot data=&DATASET;
vbox y / group=grp extreme ;
run;

title1 "Boxplots of Covariate X for Each Group";
proc sgplot data=&DATASET;
hbox x / group=grp extreme ;
```

```
run;

title1 "Scatter Plot of Response Y vs Covariate X";
proc sgplot data=&DATASET;
scatter y=y x=x / group=grp;
run;

* ANCOVA which forces the relationship between the response and the
covariate to be the same across treatment groups;

ods graphics on;

title1 "ANOVA To Compare Mean of Response Variable Y Across Groups";
proc mixed data=&DATASET;
class grp;
model y = grp ;
lsmeans grp / pdiff cl ;
run;

title1 "ANOVA To Compare Mean of Covariate X Across Groups";
proc mixed data=&DATASET;
class grp;
model x = grp ;
run;

title1 "ANCOVA To Compare Mean of Response Variable Y Across Groups,
Adjusting for X";
title2 "Parallel Slopes Model";
proc mixed data=&DATASET;
class grp;
model y = x grp ;
lsmeans grp / pdiff cl ;
run;
* Using PROC GLM to produce the ANCOVA plot, which is not produced by
PROC MIXED;

* Comment the following ODS SELECT statement out to see all of the
output from GLM;
* Leave it uncommented to output only the ANCOVA plot;

ods select ANCOVAPlot;

title1 "ANCOVA To Compare Mean of Response Variable Y Across Groups,
Adjusting for X";
title2 "Parallel Slopes Model";
proc glm data=&DATASET;
class grp;
model y = x grp ;
lsmeans grp / pdiff cl ;
run;
quit;

* ANCOVA which allows for the relationship between the response and
the covariate to differ across treatment groups. ;

ods graphics on;

title1 "ANCOVA To Compare Mean of Y Across Groups, Adjusting for X";
title2 "Non-Parallel Slopes";
proc mixed data=&DATASET;
class grp;
model y = x grp x*grp ;
lsmeans grp / pdiff cl ;
run;
* Using PROC GLM to obtain the ANCOVA plot;

* Comment the following ODS SELECT statement out to see all of the
```

```
output from GLM;
* Leave it uncommented to output only the ANCOVA plot;

ods select ANCOVAPlot;

title1 "ANCOVA To Compare Mean of Y Across Groups, Adjusting for X";
title2 "Non-Parallel Slopes";
proc glm data=&DATASET;
class grp;
model y = x grp x*grp ;
lsmeans grp / pdiff cl ;
run;
quit;

%mend;

* Choose dataset to analyze by uncommenting the appropriate line below. ;
* To uncomment a line, remove the asterisk at the beginning of the line. ;

%analyze_data(DATASET=example_1);
*%analyze_data(DATASET=example_2);
*%analyze_data(DATASET=example_3);
*%analyze_data(DATASET=example_4);

*%analyze_data(DATASET=exercise_1);
*%analyze_data(DATASET=exercise_2);
*%analyze_data(DATASET=exercise_3);
*%analyze_data(DATASET=exercise_4);
*%analyze_data(DATASET=exercise_5);
```

CHAPTER 10: ANALYSIS OF REPEATED MEASURES FOR THE BIOLOGICAL AND AGRICULTURAL SCIENCES

Salvador A. Gezan and Melissa Carvalho

Answers to True or False

1. Spatial correlation is a type of correlation that is present between observations that belong to the same experimental unit. (F)
2. If we have missing data, then repeated measures analysis can't be used. (F)
3. Combining all data from several time points into a single analysis will provide greater statistical power than analyzing every time point separately. (T)
4. For random effects, the statistical inferences are valid only for the levels that are considered in the corresponding factor. (F)
5. The compound symmetry (CS) structure is the simplest structure that can model some form of correlation. (T)
6. The AR(1) and ARH(1) structures do not need identical intervals between measurements. (F)
7. Comparing two models by using the residual log-likelihood (ReslogL) requires that the fixed effects between models are the same. (T)
8. The F- and t-tests from a repeated measures analysis are no longer valid tests because their degrees of freedom are incorrect. (F)
9. Linear mixed models can only be used on normally distributed response variables. (T)

GenStat

```
" Set working directory - change to location of your data file "
SET [WORKINGDIRECTORY='C:/…/CodeChapter']

" Read and display data from working directory "
FILEREAD [NAME='HEIGHT.TXT'; IMETHOD=read] FGROUPS=7(yes),no,no
FSPREADSHEET  Plot,Spp,Stk,Prep,Trt,Blk,Time,Initial,Ht

" Single point measurement analysis "
RESTRICT Plot,Spp,Stk,Prep,Trt,Blk,Time,Initial,Ht; CONDITION=Time.
EQ.1984
RESTRICT Plot,Spp,Stk,Prep,Trt,Blk,Time,Initial,Ht; CONDITION=Time.
EQ.1985
RESTRICT Plot,Spp,Stk,Prep,Trt,Blk,Time,Initial,Ht; CONDITION=Time.
EQ.1986
RESTRICT Plot,Spp,Stk,Prep,Trt,Blk,Time,Initial,Ht; CONDITION=Time.
EQ.1987
RESTRICT Plot,Spp,Stk,Prep,Trt,Blk,Time,Initial,Ht; CONDITION=Time.
EQ.1988
RESTRICT Plot,Spp,Stk,Prep,Trt,Blk,Time,Initial,Ht; CONDITION=Time.
```

```
EQ.1989
VCOMPONENTS [FIXED=Initial+Blk+Trt; FACTORIAL=9]
REML [PRINT=model,components,deviance,waldTests; FMETHOD=automatic;
MVINCLUDE=*; METHOD=AI;\
 MAXCYCLE=20] Ht;
VPLOT [RMETHOD=all] fittedvalues,normal,halfnormal,histogram
VPREDICT [PRINT=description,prediction,se,avesed] CLASSIFY=Trt;
LEVELS=*; PARALLEL=*
RESTRICT Plot,Spp,Stk,Prep,Trt,Blk,Time,Initial,Ht;

" Fitting different error structures with factor trt "
VCOMP [FIXED=Initial+Time+Time.Blk+Trt+Time.Trt; CADJUST=none;
FACTORIAL=9] Plot.Time;\
 CONSTRAIN=positive
VSTRUCTURE [Plot.Time] FACTOR=Time; MODEL=identity; ORDER=1;
HETEROGENEITY=none;     " ID error structure "
VSTRUCTURE [Plot.Time] FACTOR=Time; MODEL=uniform; ORDER=1;
HETEROGENEITY=none;      " CS error structure "
VSTRUCTURE [Plot.Time] FACTOR=Time; MODEL=ar; ORDER=1;
HETEROGENEITY=none;           " AR(1) error structure "
VSTRUCTURE [Plot.Time] FACTOR=Time; MODEL=diagonal; ORDER=1;
HETEROGENEITY=none;     " DIAG error structure "
VSTRUCTURE [Plot.Time] FACTOR=Time; MODEL=uniform; ORDER=1;
HETEROGENEITY=outside;   " CSH error structure "
VSTRUCTURE [Plot.Time] FACTOR=Time; MODEL=ar; ORDER=1;
HETEROGENEITY=outside;        " ARH(1) error structure "
VSTRUCTURE [Plot.Time] FACTOR=Time; MODEL=banded; ORDER=5;
HETEROGENEITY=outside;    " TOEPH error structure "
VSTRUCTURE [Plot.Time] FACTOR=Time; MODEL=unstructured; ORDER=*;
HETEROGENEITY=none; " UN error structure "
REML [PRINT=model,components,deviance,waldTests; MAXCYCLE=20;
FMETHOD=automatic; MVINCLUDE=explanatory,\
yvariate; METHOD=AI] Ht
VAIC [PRINT=deviance,aic,bic]

" Detailed model UN error structure - full treatment structure "
VCOMP [FIXED=Initial+Time+Time.Blk+Time*Spp*Stk*Prep; CADJUST=none;
FACTORIAL=17] Plot.Time;\
 CONSTRAIN=positive
VSTRUCTURE [Plot.Time] FACTOR=Time; MODEL=unstructured; ORDER=*;
HETEROGENEITY=none;  " UN error structure "
REML [PRINT=model,components,deviance,waldTests; MAXCYCLE=20;
FMETHOD=automatic; MVINCLUDE=explanatory,\
yvariate; METHOD=AI] Ht
VAIC [PRINT=deviance,aic,bic]
```

SAS

```
ods graphics on;

* Read data - change directory to location of your data file;
data HEIGHT;
  infile 'C:\...\CodeChapter\HEIGHT.TXT' firstobs=2 expandtabs;
  input Plot $ Spp $ Stk $ Prep $ Trt $ Blk $ Time $ Initial Ht;
run;
proc print data=HEIGHT (obs=20); run;

* Sorting data by Time;
proc sort data=HEIGHT;
  by Time;
run;

* Single point measurement analysis;
proc mixed data=HEIGHT plots=studentpanel;
  by Time;
  class Trt Blk Time;
  model Ht = Initial Blk Trt / ddfm=KR;
  lsmeans Trt / cl;
run;

* Fitting different error structures with factor trt;
proc mixed data=HEIGHT;
  class Plot Spp Stk Prep Trt Blk Time;
```

```
  model Ht = Initial Time Time(Blk) Trt Time*Trt / ddfm=KR;
  *repeated Time / subject=Plot type=VC;            * ID error
structure;
  *repeated Time / subject=Plot type=CS;            * CS error
structure;
  *repeated Time / subject=Plot type=AR(1);         * AR(1) error
structure;
  *repeated Time / subject=Plot type=TOEP;       * TOEP error structure;
  *repeated Time / subject=Plot type=VC group=Time; * DIAG error structure;
  *repeated Time / subject=Plot type=CSH;         * CSH error structure;
  *repeated Time / subject=Plot type=ARH(1);     * ARH(1) error structure;
  *repeated Time / subject=Plot type=TOEPH;      * TOEPH error structure;
  repeated Time / subject=Plot type=UN;          * UN error structure;
  lsmeans Time*Trt / slice=Time;
run;

* Detailed model UN error structure - full treatment structure;
proc mixed data=HEIGHT plots=studentpanel;
  class Plot Spp Stk Prep Trt Blk Time;
  model Ht = Initial Time Time(Blk) Spp|Stk|Prep Time*Spp Time*Stk
Time*Prep Time*Spp*Stk Time*Spp*Prep Time*Stk*Prep Time*Spp*Stk*Prep /
ddfm=KR residual outp=resid;
  repeated Time / subject=Plot type=UN r rcorr;
run;

ods graphics off;
```

R

```
# Read data - change directory to location of your data file;
rm(list=ls())
setwd("C:/Users/sgeza/Desktop/Repeatead/Paper_AGJournal_2016/
Revisions_Nov2016")

# Loading libraries
library(nlme)
library(lsmeans)

# Reading data
datasoy<-data.frame(Soybean[Soybean$Year=='1988',])
head(datasoy)

# Defining factors
datasoy$Plot<-as.factor(datasoy$Plot)
datasoy$Variety<-as.factor(datasoy$Variety)
datasoy$Time<-as.factor(datasoy$Time)
str(datasoy)

# Some EDA
boxplot(weight~Time,data=datasoy)
hist(datasoy$weight)
datasoy$logweight<-log(datasoy$weight)
boxplot(logweight~Time,data=datasoy)

# Obtaining subsets of the data by Time
T14<-datasoy[datasoy$Time==14,]

# Single point measurement analysis
modelSingle<-lm(logweight~Variety,data=T14)
summary(modelSingle)
anova(modelSingle)
plot(modelSingle)
lsmeans(modelSingle,~Variety)

# ID error structure
rstruct<-varIdent(form=~1)
rheter<-varIdent(form=~1)
modID<-gls(logweight~Time+Variety+Time:Variety,
           correlation=rstruct,weights=rheter,data=datasoy)
#summary(modID)
#anova(modID)
#plot(modID)
```

```
# CS error structure
rstruct<-corCompSymm(form=~Time|Plot)
rheter<-varIdent(form=~1)
modCS<-gls(logweight~Time+Variety+Time:Variety,
           correlation=rstruct,weights=rheter,data=datasoy)

# CSH error structure
rstruct<-corCompSymm(form=~Time|Plot)
rheter<-varIdent(form=~1|Time)
modCSH<-gls(logweight~Time+Variety+Time:Variety,
           correlation=rstruct,weights=rheter,data=datasoy)

# DIAG error structure
rstruct<-varIdent(form=~1)
rheter<-varIdent(form=~1|Time)
modDIAG<-gls(logweight~Time+Variety+Time:Variety,
           correlation=rstruct,weights=rheter,data=datasoy)

# AR(1) error structure
rstruct<-corAR1(form=~1|Plot)
rheter<-varIdent(form=~1)
modAR1<-gls(logweight~Time+Variety+Time:Variety,
           correlation=rstruct,weights=rheter,data=datasoy)

# ARH(1) error structure
rstruct<-corAR1(form=~1|Plot)
rheter<-varIdent(form=~1|Time)
modARH1<-gls(logweight~Time+Variety+Time:Variety,
           correlation=rstruct,weights=rheter,data=datasoy)

# US error structure (with extra output)
rstruct<-corSymm(form=~1|Plot)
rheter<-varIdent(form=~1|Time)
modUS<-gls(logweight~Time+Variety+Time:Variety,
           correlation=rstruct,weights=rheter,data=datasoy)

# Comparing US against ARH(1) using Likelihood ratio test
anova(modUS,modARH1)

# Output for selected model
#mod<-modID
#mod<-modCS
#mod<-modCSH
#mod<-modDIAG
#mod<-modAR1
mod<-modARH1
#mod<-modUS
output<-summary(mod)
anova(mod,type='marginal')   # Marginal ANOVA table
anova(mod,type='sequential') # Sequential ANOVA table
attr(output$apVar,"Pars")    # Variance components
(logL<-2*mod$logLik)         # log-likelihood value
(AIC<-output$AIC)            # AIC
(BIC<-output$BIC)            # BIC
```

CHAPTER 11: THE DESIGN AND ANALYSIS OF LONG-TERM ROTATION EXPERIMENTS

Roger William Payne

Appendix 1. Genstat commands to analyze the potato yields from the Westmaas experiment.

```
IMPORT [PRINT=*] 'wmpotato.xlsx'
" try various random models "
CAPTION 'Split-plot nested within years'; STYLE=meta
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen] Year/Block/Wholeplot
REML [PRINT=components] Yield
VAIC [PRINT=deviance,aic,sic,dfrandom]
VRACCUMULATE [PRINT=*] 'Split-plot nested within years'
CAPTION 'Nested split-plot, different residual variance each year';\
  STYLE=meta
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen; EXPERIMENT=Year]\
  Year/Block/Wholeplot
REML [PRINT=components] Yield
VAIC [PRINT=deviance,aic,sic,dfrandom]
VRACCUMULATE [PRINT=*] 'Nested split-plot meta analysis'
" Nested split plot: power-distance correlation structure over years"
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen] Year/Block/Wholeplot
VARIATE Ycoord; VALUES=Year
VSTRUCTURE [TERM=Year.Block.Wholeplot; COORDINATES=Ycoord] power;\
  FACTOR=Year
REML [PRINT=*] Yield
VRACCUMULATE [PRINT=*] 'Nested split-plot and power distance'
" Nested split plot with different residual variance in each year
  and power-distance correlation structure over years "
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen; EXPERIMENT=Year]\
  Year/Block/Wholeplot
VSTRUCTURE [TERM=Year.Block.Wholeplot; COORDINATES=Ycoord] power;\
  FACTOR=Year
REML [PRINT=*] Yield
VRACCUMULATE [PRINT=*]\
  'Nested split-plot meta analysis and power distance'
" Nested split plot with different residual variances
  and variance components "
FORMULA [VALUE=Block/Wholeplot] differentvcterms
VRMETA [EXPERIMENTSFACTOR=Year; RANDOM=Random] 77,79,80...88;\
  LOCALTERMS=differentvcterms
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen; EXPERIMENTS=Year] #Random
REML [PRINT=*; MVINCLUDE=explanatory] Yield
VRACCUMULATE 'Meta analysis with different variance components'
" use split plot with different residual variance in each year "
CAPTION 'Nested split-plot, different residual variance each year';\
  STYLE=meta
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen; EXPERIMENT=Year]\
  Year/Block/Wholeplot
REML [PRINT=wald] Yield
" drop unnecessary fixed terms "
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen-Year.Rotation.Nitrogen;\
  EXPERIMENTS=Year] Year/Block/Wholeplot
REML [PRINT=wald; MVINCLUDE=explanatory; WORKSPACE=100] Yield
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen - Year.Rotation.Nitrogen\
  - Rotation.Nitrogen; EXPERIMENTS=Year] Year/Block/Wholeplot
REML [PRINT=wald; MVINCLUDE=explanatory; WORKSPACE=100] Yield
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen - Year.Rotation.Nitrogen\
  - Rotation.Nitrogen - Year.Nitrogen; EXPERIMENTS=Year]\
```

```
  Year.Block/Wholeplot
REML [PRINT=wald; MVINCLUDE=explanatory; WORKSPACE=100] Yield
" predicted means for Nitrogen,
  and for Year.Rotation with a summary of the sed's "
VDISPLAY [PRINT=means; PTERMS=Nitrogen+Year.Rotation]
" to print the means with all the sed's, set option PSE
VDISPLAY [PRINT=means; PTERMS=Year.Rotation; PSE=alldifferences]"
PEN 11...15; LINESTYLE=1,8,1,8,1;\
  CSYMBOL='crimson','darkviolet','violet','darkblue','royalblue';\
  CLINE='crimson','darkviolet','violet','darkblue','royalblue';\
  CFILL='crimson','darkviolet','violet','darkblue','royalblue';\
  SYMBOL='circle','heavycross','diamond','heavyplus','square'
VGRAPH [METHOD=linesandpoints] Year; GROUP=Rotation;\
  PENS=!(11...15); YTITLE='Yield (t ha~^{-1})';\
  TITLE='Estimated mean yield of potato for rotations and years'
```

Appendix 2. Genstat output for the potato yields from the Westmaas experiment.

```
   2  IMPORT [PRINT=*] 'wmpotato.xlsx'
   3  " try various random models "
   4  CAPTION 'Split-plot nested within years'; STYLE=meta
```

Split-plot nested within years

```
   5  VCOMPONENTS [FIXED=Year*Rotation*Nitrogen] Year/Block/Wholeplot
   6  REML [PRINT=components] Yield
```

Estimated variance components

Random Term	Component	S.e.
Year.Block	1.407	1.403
Year.Block.Wholeplot	7.079	1.893

Residual variance model

Term	Model(order)	Parameter	Estimate	s.e.
Residual	Identity	Sigma2	5.195	0.701

```
   7  VAIC [PRINT=deviance,aic,sic,dfrandom]
```

Deviance	950.43
Akaike information coefficient	956.43
Schwarz Bayes information coefficient	965.75
d.f. of random model	3

Note: omits constant, -log(det(X'X)), that depends only on the fixed model.

```
   8  VRACCUMULATE [PRINT=*] 'Split-plot nested within years'
   9  CAPTION 'Nested split-plot, different residual variance each year';\
  10     STYLE=meta
```

Nested split-plot, different residual variance each year

```
  11  VCOMPONENTS [FIXED=Year*Rotation*Nitrogen; EXPERIMENT=Year]\
  12     Year/Block/Wholeplot
  13  REML [PRINT=components] Yield
```

Estimated variance components

Random term	component	s.e.
Year.Block	2.12	1.17
Year.Block.Wholeplot	0.91	0.52

Residual model for each experiment
Experiment factor: Year

Experiment	Term	Factor	Model(order)	Parameter	Estimate	s.e.
77.00	Residual		Identity	Variance	0.819	0.352
79.00	Residual		Identity	Variance	2.684	1.149
80.00	Residual		Identity	Variance	1.344	0.588
81.00	Residual		Identity	Variance	5.231	2.558
82.00	Residual		Identity	Variance	4.538	1.873
83.00	Residual		Identity	Variance	1.988	0.887
84.00	Residual		Identity	Variance	4.013	1.660
85.00	Residual		Identity	Variance	5.415	2.185
86.00	Residual		Identity	Variance	14.28	5.56
87.00	Residual		Identity	Variance	58.21	22.43
88.00	Residual		Identity	Variance	12.90	5.25

```
VAIC [PRINT=deviance,aic,sic,dfrandom]
```

```
Deviance                                   891.60
Akaike information coefficient             917.60
Schwarz Bayes information coefficient      957.98
d.f. of random model                       13
```

Note: omits constant, -log(det(X'X)), that depends only on the fixed model.

```
VRACCUMULATE [PRINT=*] 'Nested split-plot meta analysis'
" Nested split plot: power-distance correlation structure over years"
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen] Year/Block/Wholeplot
VARIATE Ycoord; VALUES=Year
VSTRUCTURE [TERM=Year.Block.Wholeplot; COORDINATES=Ycoord] power;\
   FACTOR=Year
REML [PRINT=*] Yield
VRACCUMULATE [PRINT=*] 'Nested split-plot and power distance'
" Nested split plot with different residual variance in each year
   and power-distance correlation structure over years "
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen; EXPERIMENT=Year]\
   Year/Block/Wholeplot
VSTRUCTURE [TERM=Year.Block.Wholeplot; COORDINATES=Ycoord] power;\
   FACTOR=Year
REML [PRINT=*] Yield
VRACCUMULATE [PRINT=*]\
   'Nested split-plot meta analysis and power distance'
" Nested split plot with different residual variances
   and variance components "
FORMULA [VALUE=Block/Wholeplot] differentvcterms
VRMETA [EXPERIMENTSFACTOR=Year; RANDOM=Random] 77,79,80...88;\
   LOCALTERMS=differentvcterms
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen; EXPERIMENTS=Year] #Random
REML [PRINT=*; MVINCLUDE=explanatory] Yield
VRACCUMULATE 'Meta analysis with different variance components'
```

Accumulated summary of REML random models

	Deviance	AICSIC		Random d.f.
Split-plot nested within years	950.43	956.43	965.75	3
Nested split-plot meta analysis	891.60	917.60	957.98	13
Nested split-plot and power distance	950.38	958.38	970.80	4
Nested split-plot meta analysis and power distance	891.57	919.57	963.06	14
Meta analysis with different variance components	852.53	918.53	1021.02	33

Note: omits constant, -log(det(X'X)), that depends only on the fixed model.

```
" use split plot with different residual variance in each year "
CAPTION 'Nested split-plot, different residual variance each year';\
   STYLE=meta
```

Nested split-plot, different residual variance each year

```
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen; EXPERIMENT=Year]\
   Year/Block/Wholeplot
REML [PRINT=wald] Yield
```

Tests for fixed effects

Sequentially adding terms to fixed model

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
Year	851.65	10	85.45	10.4	<0.001
Rotation	106.93	4	26.73	27.0	<0.001
Nitrogen	110.56	2	55.28	58.1	<0.001
Year.Rotation	154.93	40	3.75	22.8	<0.001
Year.Nitrogen	29.16	20	1.27	48.7	0.242
Rotation.Nitrogen	15.43	8	1.93	58.1	0.073
Year.Rotation.Nitrogen	108.45	80	1.19	40.6	0.271

Dropping individual terms from full fixed model

Fixed term	Wald statistic	n.d.f	F statistic	d.d.f.	F pr
Year.Rotation.Nitrogen	108.45	80	1.19	40.6	0.271

Message: denominator degrees of freedom for approximate F-tests are calculated using algebraic derivatives ignoring fixed/boundary/singular variance parameters.

```
" drop unnecessary fixed terms "
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen-Year.Rotation.Nitrogen;\
   EXPERIMENTS=Year] Year/Block/Wholeplot
REML [PRINT=wald; MVINCLUDE=explanatory; WORKSPACE=100] Yield
```

Tests for Fixed Effects

Sequentially adding terms to fixed model

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
Year	796.16	10	79.79	10.7	<0.001
Rotation	91.94	4	22.98	32.8	<0.001
Nitrogen	83.70	2	41.85	128.4	<0.001
Year.Rotation	117.01	40	2.92	25.3	0.003
Year.Nitrogen	29.67	20	1.36	88.3	0.166
Rotation.Nitrogen	10.34	8	1.29	128.4	0.253

Dropping individual terms from full fixed model

Fixed term	Wald statistic	n.d.f.	*F* statistic	d.d.f.	F pr
Year.Rotation	117.01	40	2.92	25.3	0.003
Year.Nitrogen	29.67	20	1.36	88.3	0.166
Rotation.Nitrogen	10.34	8	1.29	128.4	0.253

Message: denominator degrees of freedom for approximate *F*-tests are calculated using algebraic derivatives ignoring fixed/boundary/singular variance parameters.

```
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen - Year.Rotation.Nitrogen\
   - Rotation.Nitrogen; EXPERIMENTS=Year] Year/Block/Wholeplot
REML [PRINT=wald; MVINCLUDE=explanatory; WORKSPACE=100] Yield
```

Tests for fixed effects

Sequentially adding terms to fixed model

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
Year	783.23	10	78.49	10.8	<0.001
Rotation	89.66	4	22.42	34.0	<0.001
Nitrogen	83.72	2	41.86	136.7	<0.001
Year.Rotation	112.76	40	2.82	26.9	0.003
Year.Nitrogen	29.82	20	1.38	95.6	0.154

Dropping individual terms from full fixed model

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
Year.Rotation	112.76	40	2.82	26.9	0.003
Year.Nitrogen	29.82	20	1.38	95.6	0.154

Message: denominator degrees of freedom for approximate F-tests are calculated using algebraic derivatives ignoring fixed/boundary/singular variance parameters.

```
VCOMPONENTS [FIXED=Year*Rotation*Nitrogen - Year.Rotation.Nitrogen\
   - Rotation.Nitrogen - Year.Nitrogen; EXPERIMENTS=Year]\
   Year.Block/Wholeplot
REML [PRINT=wald; MVINCLUDE=explanatory; WORKSPACE=100] Yield
```

Tests for fixed effects

Sequentially adding terms to fixed model

Fixed term	Wald statistic	n.d.f.	*F* statistic	d.d.f.	F pr
Year	780.43	10	78.21	10.8	<0.001
Rotation	92.68	4	23.17	32.6	<0.001
Nitrogen	83.21	2	41.60	142.0	<0.001
Year.Rotation	117.47	40	2.93	27.0	0.002

Dropping individual terms from full fixed model

Fixed term	Wald statistic	n.d.f.	*F* statistic	d.d.f.	F pr
Nitrogen	83.21	2	41.60	142.0	<0.001
Year.Rotation	117.47	40	2.93	27.0	0.002

Message: denominator degrees of freedom for approximate F-tests are calculated using algebraic derivatives ignoring fixed/boundary/singular variance parameters.

```
" predicted means for Nitrogen,
  and for Year.Rotation with a summary of the sed's "
VDISPLAY [PRINT=means; PTERMS=Nitrogen+Year.Rotation]
```

Table of predicted means for Nitrogen

Nitrogen	N1	N2	N3
	43.37	45.03	45.65

Standard error of differences: 0.2585

Table of predicted means for Year.Rotation

Rotation Year	IIf	III	IIIf	IV	IVf
77	23.44	21.92	21.87	25.01	21.20
79	39.44	46.62	45.74	49.50	44.35
80	40.53	41.14	40.90	41.54	42.48
81	30.37	38.04	39.45	39.43	38.14
82	39.75	40.49	42.95	45.53	45.63
83	31.14	36.55	39.53	37.96	37.61
84	54.59	57.67	56.63	58.40	58.99
85	51.48	47.80	49.79	52.54	49.75
86	48.21	51.03	54.58	51.48	52.99
87	42.99	48.76	41.56	45.67	51.85
88	54.28	62.76	64.41	58.52	62.46

Standard errors of differences

Average	2.483
Maximum	3.686
Minimum	1.413

Average variance of differences: 6.366

Standard error of differences for same level of factor:

```
                Year    Rotation
Average:        1.978   2.524
Maximum:        3.686   3.527
Minimum:        1.413   2.074

Average variance of differences:
                4.328   6.529
```

```
" to print the means with all the sed's, set option PSE
VDISPLAY [PRINT=means; PTERMS=Year.Rotation; PSE=alldifferences]"
PEN 11...15; LINESTYLE=1,8,1,8,1;\
 CSYMBOL='crimson','darkviolet','violet','darkblue','royalblue';\
 CLINE='crimson','darkviolet','violet','darkblue','royalblue';\
 CFILL='crimson','darkviolet','violet','darkblue','royalblue';\
 SYMBOL='circle','heavycross','diamond','heavyplus','square'
VGRAPH [METHOD=linesandpoints] Year; GROUP=Rotation;\
  PENS=!(11...15); YTITLE='Yield (t ha~^{-1})';\
  TITLE='Estimated mean yield of potato for rotations and years'
```

Appendix 3. ASReml-R commands to analyze potato yields from the Westmaas experiment.

```
#'## Initialize
library(asreml4)
library(asremlPlus)
library(ggplot2)
library(knitr)
#knitr::spin("Appendix1.v4.r")
options(width = 110)

#'## Load data, order it and create a 3 level nested Wholeplot factor
load(file = "wmpotato.rda")
```

```
wmpotato <- with(wmpotato, wmpotato[order(Year,Block,Wholeplot,Subplot), ])
wmpotato$NWholeplot <- factor(rep(1:5, each=3, times=22))

#'## Try various random models
#'### Split plot within years
model1.asr <- asreml(Yield ~ Year*Rotation*Nitrogen,
                      random = ~Year:Block/Wholeplot,
                      data = wmpotato)
summary(model1.asr)$varcomp
info.accumulate <- data.frame(Model = "Split-plot",
                              infoCriteria(model1.asr),
                              stringsAsFactors = FALSE)

#'### Split plot with different residual variance in each year
model2.asr <- asreml(Yield ~ Year*Rotation*Nitrogen,
                     random = ~Year:Block/Wholeplot,
                     residual = ~ idh(Year):Block:NWholeplot:Subplot,
                     data = wmpotato)
model2.asr <- update(model2.asr)
summary(model2.asr)$varcomp
info.accumulate <- rbind(info.accumulate,
                         data.frame(Model = "Split plot meta analysis",
                                    infoCriteria(model2.asr),
                                    stringsAsFactors = FALSE))

#'### Split plot with EXP structure over years
model3.asr <- asreml(Yield ~ Year*Rotation*Nitrogen,
                     random = ~Year:Block/Wholeplot,
                     residual = ~ exp(Year):Block:NWholeplot:Subplot,
                     data = wmpotato)
summary(model3.asr)$varcomp
info.accumulate <- rbind(info.accumulate,
                         data.frame(Model = "Split-plot and EXP",
                                    infoCriteria(model3.asr),
                                    stringsAsFactors = FALSE))

#'### Split plot: different residual variance in each year & EXP
structure
model4.asr <- asreml(Yield ~ Year*Rotation*Nitrogen,
                    random = ~Year:Block/Wholeplot,
                    residual = ~ exph(Year):Block:NWholeplot:Subplot,
                     data = wmpotato)
model4.asr <- update(model4.asr)
summary(model4.asr)$varcomp
info.accumulate <- rbind(info.accumulate,
                   data.frame(Model = "Split-plot meta analysis and EXP",
                                    infoCriteria(model4.asr),
                                    stringsAsFactors = FALSE))

dimnames(summary(model4.asr)$varcomp)

#'### Split plot with different residual variances and variance
components
model5.asr <- asreml(Yield ~ Year*Rotation*Nitrogen,
                     random = ~ idh(Year):Block/Wholeplot,
                     residual =   ~ idh(Year):Block:NWholeplot:Subplot,
                     data = wmpotato)
summary(model5.asr)$varcomp
vcnames <- rownames(summary(model5.asr)$varcomp)[1:22]
model5.asr <- setvarianceterms(model5.asr$call, terms = vcnames,
                   bounds = "U", initial=0.01, ignore.suffices = FALSE)
model5.asr <- update(model5.asr)
summary(model5.asr)$varcomp
info.accumulate <- rbind(info.accumulate,
     data.frame(Model = "Meta analysis with different variance components",
                          infoCriteria(model5.asr, bound.exclusions = "F"),
                           stringsAsFactors = FALSE))
```

```
#'### Accumulated summary of REML information criteria
info.accumulate
#'### Use split plot with different residual variance in each year
wald(model2.asr, denDF = "algebraic")

#'## Drop unnecessary fixed terms
model2a.asr <- asreml(Yield ~ Year*Rotation*Nitrogen-Year:Rotation:Nitrogen,
                      random = ~ Year:Block/Wholeplot,
                     residual = ~ idh(Year):Block:NWholeplot:Subplot,
                      data = wmpotato)
wald(model2a.asr, denDF = "algebraic")

model2b.asr <- asreml(Yield ~ Year*Rotation*Nitrogen-
Year:Rotation:Nitrogen -
                                          Rotation:Nitrogen,
                       random = ~ Year:Block/Wholeplot,
                     residual =   ~ idh(Year):Block:NWholeplot:Subplot,
                       data = wmpotato)
wald(model2b.asr, denDF = "algebraic")

model2c.asr <- asreml(Yield ~ Year*Rotation*Nitrogen-
Year:Rotation:Nitrogen -
                         Rotation:Nitrogen - Year:Nitrogen,
                       random = ~ Year:Block/Wholeplot,
                       residual =   ~ idh(Year):Block:NWholeplot:Subplot,
                       data = wmpotato)
wald(model2c.asr, denDF = "algebraic")

#'## Get predictions and plot
predict(model2c.asr, classify = "Nitrogen")$pvals
predYR <- predict(model2c.asr, classify = "Year:Rotation")$pvals
predYR

cols <- c('red','darkviolet','violet','darkblue','lightskyblue')
ggplot(data = predYR,
      aes(x =Year, y=predicted.value, colour=Rotation, linetype=Rotation,
       shape = Rotation)) + geom_point() + geom_line() +
       labs(y = "Yield") + scale_color_manual(values =
       cols) + scale_shape_manual(values = c(16,4,18,3,15))
```

Appendix 4. ASReml-R output from the analysis of potato yields from the Westmaas experiment.

Initialize

```
library(asreml4)
## Loading required package: Matrix
## Licensed to University of South Australia, serial number 402060331,
expires 31-jan-2018, 373 days.
library(asremlPlus)
library(ggplot2)
library(knitr)
#knitr::spin("Appendix1.v4.r")
options(width = 110)
```

Load data, order it and create a 3 level nested Wholeplot factor

```
load(file = "wmpotato.rda")
wmpotato <- with(wmpotato, wmpotato[order(Year,Block,Wholeplot,Subplot), ])
wmpotato$NWholeplot <- factor(rep(1:5, each=3, times=22))
```

Try various random models

Split plot within years

```
model1.asr <- asreml(Yield ~ Year*Rotation*Nitrogen,
                     random = ~Year:Block/Wholeplot,
                     data = wmpotato)
```

```
## Model fitted using the gamma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:32 2017
##           LogLik        Sigma2     DF     wall     cpu
##  1      -341.227       10.0174    165 15:55:32    0.0
##  2      -332.725        7.9880    165 15:55:32    0.0
##  3      -326.355        6.4141    165 15:55:32    0.0
##  4      -324.012        5.5986    165 15:55:32    0.0
##  5      -323.596        5.2382    165 15:55:32    0.0
##  6      -323.590        5.1960    165 15:55:32    0.0
##  7      -323.590        5.1952    165 15:55:32    0.0
summary(model1.asr)$varcomp
##                       component std.error  z.ratio bound %ch
## Year:Block             1.407175 1.4026645 1.003216     P   0
## Year:Block:Wholeplot   7.079126 1.8929384 3.739755     P   0
## units(R)               5.195198 0.7005203 7.416198     P   0
info.accumulate <- data.frame(Model = "Split-plot",
                              infoCriteria(model1.asr),
                              stringsAsFactors = FALSE)
Split plot with different residual variance in each year
model2.asr <- asreml(Yield ~ Year*Rotation*Nitrogen,
                     random = ~Year:Block/Wholeplot,
                     residual = ~ idh(Year):Block:NWholeplot:Subplot,
                     data = wmpotato)
## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:32 2017
##           LogLik        Sigma2     DF     wall     cpu
##  1      -325.726          1.0    165 15:55:32    0.0 (2 restrained)
##  2      -315.906          1.0    165 15:55:32    0.0
##  3      -307.104          1.0    165 15:55:32    0.0
##  4      -300.698          1.0    165 15:55:32    0.0
##  5      -295.773          1.0    165 15:55:32    0.0
##  6      -294.551          1.0    165 15:55:32    0.0
##  7      -294.313          1.0    165 15:55:32    0.0
##  8      -294.249          1.0    165 15:55:32    0.0
##  9      -294.216          1.0    165 15:55:32    0.0
## 10      -294.197          1.0    165 15:55:32    0.0
## 11      -294.187          1.0    165 15:55:32    0.0
## 12      -294.181          1.0    165 15:55:32    0.0
## 13      -294.178          1.0    165 15:55:32    0.0
## Warning in asreml(Yield ~ Year * Rotation * Nitrogen, random =
~Year:Block/Wholeplot, : Log-likelihood not
## converged
## Warning in asreml(Yield ~ Year * Rotation * Nitrogen, random =
~Year:Block/Wholeplot, : Some components
## changed by more than 1% on the last iteration.
model2.asr <- update(model2.asr)
## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:32 2017
##           LogLik        Sigma2     DF     wall     cpu
##  1      -294.176          1.0    165 15:55:32    0.0
##  2      -294.176          1.0    165 15:55:32    0.0
##  3      -294.176          1.0    165 15:55:32    0.0
##  4      -294.175          1.0    165 15:55:32    0.0
##  5      -294.175          1.0    165 15:55:32    0.0
summary(model2.asr)$varcomp
##                                        component  std.error  z.ratio bound %ch
## Year:Block                             2.1188142  1.1698023 1.811258    P 0.0
## Year:Block:Wholeplot                   0.9225528  0.5212748 1.769801    P 0.5
## Year:Block:NWholeplot:Subplot(R)       1.0000000         NA       NA    F 0.0
## Year:Block:NWholeplot:Subplot!Year_77  0.8192466  0.3522389 2.325826    P 0.0
## Year:Block:NWholeplot:Subplot!Year_79  2.6812820  1.1480199 2.335571    P 0.0
## Year:Block:NWholeplot:Subplot!Year_80  1.3431204  0.5869084 2.288467    P 0.0
## Year:Block:NWholeplot:Subplot!Year_81  5.1776315  2.5260176 2.049721    P 0.5
## Year:Block:NWholeplot:Subplot!Year_82  4.5354571  1.8726148 2.421991    P 0.0
## Year:Block:NWholeplot:Subplot!Year_83  1.9832480  0.8836502 2.244381    P 0.1
## Year:Block:NWholeplot:Subplot!Year_84  4.0128598  1.6612209 2.415609    P 0.0
## Year:Block:NWholeplot:Subplot!Year_85  5.4153625  2.1867447 2.476449    P 0.0
## Year:Block:NWholeplot:Subplot!Year_86 14.2810395  5.5608347 2.568147    P 0.0
## Year:Block:NWholeplot:Subplot!Year_87 58.1502751 22.4010672 2.595871    P 0.0
## Year:Block:NWholeplot:Subplot!Year_88 12.8700511  5.2352869 2.458328    P 0.1
info.accumulate <- rbind(info.accumulate,
                         data.frame(Model = "Split plot meta analysis",
                                    infoCriteria(model2.asr),
                                    stringsAsFactors = FALSE))
## Warning in infoCriteria.asreml(model2.asr): The following bound terms
```

```
were discounted:
##  Year:Block:NWholeplot:Subplot(R)
Split plot with EXP structure over years
model3.asr <- asreml(Yield ~ Year*Rotation*Nitrogen,
                     random = ~Year:Block/Wholeplot,
                    residual = ~ exp(Year):Block:NWholeplot:Subplot,
                     data = wmpotato)
## Model fitted using the gamma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:32 2017
##          LogLik      Sigma2    DF     wall    cpu
##  1     -340.831     10.0380    165 15:55:32   0.0
##  2     -332.931      8.0998    165 15:55:32   0.0
##  3     -326.549      6.4773    165 15:55:32   0.0
##  4     -324.050      5.6183    165 15:55:32   0.0
##  5     -323.597      5.2428    165 15:55:32   0.0
##  6     -323.590      5.1963    165 15:55:32   0.0
##  7     -323.590      5.1952    165 15:55:32   0.0
## Warning in asreml(Yield ~ Year * Rotation * Nitrogen, random =
~Year:Block/Wholeplot, : Some components
## changed by more than 1% on the last iteration.
summary(model3.asr)$varcomp
##                    component std.error    z.ratio bound  %ch
## Year:Block        1.407182e+00 1.4026520 1.0032293744     P  0.0
## Year:Block:Wholeplot      7.079156e+00 1.8933937 3.7388719187     P  0.0
## Year:Block:NWholeplot:Subplot(R)     5.195228e+00 0.7006187 7.4151999388   P  0.0
## Year:Block:NWholeplot:Subplot!Year!pow 1.090783e-05 0.1010570 0.0001079374   U 58.6
info.accumulate <- rbind(info.accumulate,
                         data.frame(Model = "Split-plot and EXP",
                                    infoCriteria(model3.asr),
                                    stringsAsFactors = FALSE))
Split plot with different residual variance in each year and EXP structure
model4.asr <- asreml(Yield ~ Year*Rotation*Nitrogen,
                     random = ~Year:Block/Wholeplot,
                    residual = ~ exph(Year):Block:NWholeplot:Subplot,
                     data = wmpotato)
## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:32 2017
##          LogLik       Sigma2     DF     wall     cpu
##  1    -326.106          1.0   165 15:55:32    0.0 (2 restrained)
##  2     -315.850          1.0    165 15:55:32    0.0
##  3     -307.051          1.0    165 15:55:32    0.0
##  4     -300.617          1.0    165 15:55:32    0.0
##  5     -295.655          1.0    165 15:55:32    0.0
##  6     -294.412          1.0    165 15:55:32    0.0
##  7     -294.167          1.0    165 15:55:32    0.0
##  8     -294.097          1.0    165 15:55:32    0.0
##  9     -294.060          1.0    165 15:55:32    0.0
## 10     -294.037          1.0    165 15:55:32    0.0
## 11     -294.024          1.0    165 15:55:32    0.0
## 12     -294.016          1.0    165 15:55:32    0.0
## 13     -294.012          1.0    165 15:55:32    0.0
## Warning in asreml(Yield ~ Year * Rotation * Nitrogen, random =
~Year:Block/Wholeplot, : Log-likelihood not
## converged
## Warning in asreml(Yield ~ Year * Rotation * Nitrogen, random =
~Year:Block/Wholeplot, : Some components
## changed by more than 1% on the last iteration.
model4.asr <- update(model4.asr)
## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:32 2017
##          LogLik       Sigma2     DF     wall     cpu
##  1     -294.009          1.0    165 15:55:32    0.0
##  2     -294.009          1.0    165 15:55:33    0.0
##  3     -294.008          1.0    165 15:55:33    0.0
##  4     -294.007          1.0    165 15:55:33    0.0
##  5     -294.007          1.0    165 15:55:33    0.0
##  6     -294.007          1.0    165 15:55:33    0.0
summary(model4.asr)$varcomp
##                 component  std.error   z.ratio bound %ch
## Year:Block                            2.12260198  1.1684830 1.8165450
P 0.0
## Year:Block:Wholeplot                0.84472627  0.5064265  1.6680136    P 0.6
## Year:Block:NWholeplot:Subplot(R)    1.00000000     NA       NA         F 0.0
## Year:Block:NWholeplot:Subplot!Year!pow  0.06584539  0.1081466 0.6088532    U 0.2
## Year:Block:NWholeplot:Subplot!Year_77    0.81429437  0.3493945 2.3305874     P 0.0
```

```
## Year:Block:NWholeplot:Subplot!Year_79   2.68405000  1.1469765 2.3401091     P 0.0
## Year:Block:NWholeplot:Subplot!Year_80   1.34566016  0.5892322 2.2837521     P 0.0
## Year:Block:NWholeplot:Subplot!Year_81   5.19119457  2.5411514 2.0428513     P 0.5
## Year:Block:NWholeplot:Subplot!Year_82   4.51796149  1.8574883 2.4322961     P 0.0
## Year:Block:NWholeplot:Subplot!Year_83   2.11597580  0.9675121 2.1870277     P 0.1
## Year:Block:NWholeplot:Subplot!Year_84   4.13956058  1.7259741 2.3983910     P 0.0
## Year:Block:NWholeplot:Subplot!Year_85   5.48604076  2.2146976 2.4771060     P 0.0
## Year:Block:NWholeplot:Subplot!Year_86 14.41228320 5.5903544 2.5780625  P 0.0
## Year:Block:NWholeplot:Subplot!Year_87  58.23033685  22.3474570   2.6056807   P 0.0
## Year:Block:NWholeplot:Subplot!Year_88  13.12521093  5.3138979 2.4699780     P 0.1
info.accumulate <- rbind(info.accumulate,
                         data.frame(Model = "Split-plot meta analysis and EXP",
                                    infoCriteria(model4.asr),
                                    stringsAsFactors = FALSE))
## Warning in infoCriteria.asreml(model4.asr): The following bound terms
were discounted:
##   Year:Block:NWholeplot:Subplot(R)
dimnames(summary(model4.asr)$varcomp)
## [[1]]
##  [1] "Year:Block"                              "Year:Block:Wholeplot"
##  [3] "Year:Block:NWholeplot:Subplot(R)"
"Year:Block:NWholeplot:Subplot!Year!pow"
##  [5] "Year:Block:NWholeplot:Subplot!Year_77"
"Year:Block:NWholeplot:Subplot!Year_79"
##  [7] "Year:Block:NWholeplot:Subplot!Year_80"
"Year:Block:NWholeplot:Subplot!Year_81"
##  [9] "Year:Block:NWholeplot:Subplot!Year_82"
"Year:Block:NWholeplot:Subplot!Year_83"
## [11] "Year:Block:NWholeplot:Subplot!Year_84"
"Year:Block:NWholeplot:Subplot!Year_85"
## [13] "Year:Block:NWholeplot:Subplot!Year_86"
"Year:Block:NWholeplot:Subplot!Year_87"
## [15] "Year:Block:NWholeplot:Subplot!Year_88"
##
## [[2]]
## [1] "component" "std.error" "z.ratio"   "bound"     "%ch"
Split plot with different residual variances and variance components
model5.asr <- asreml(Yield ~ Year*Rotation*Nitrogen,
                     random = ~ idh(Year):Block/Wholeplot,
                   residual =   ~ idh(Year):Block:NWholeplot:Subplot,
                     data = wmpotato)
## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:33 2017
##          LogLik        Sigma2      DF      wall      cpu
##  1      -325.726          1.0     165 15:55:33    0.0 (13 restrained)
##  2      -298.707          1.0     165 15:55:33    0.0 (6 restrained)
##  3      -285.216          1.0     165 15:55:33    0.0 (6 restrained)
##  4      -278.707          1.0     165 15:55:33    0.0 (7 restrained)
##  5      -276.570          1.0     165 15:55:33    0.0 (8 restrained)
##  6      -275.950          1.0     165 15:55:33    0.0 (5 restrained)
##  7      -275.800          1.0     165 15:55:33    0.0 (1 restrained)
##  8      -275.795          1.0     165 15:55:33    0.0 (1 restrained)
##  9      -275.795          1.0     165 15:55:33    0.0
## 10      -275.795          1.0     165 15:55:33    0.0
summary(model5.asr)$varcomp
##                         component  std.error    z.ratio bound %ch
## Year:Block!Year_7    2.321580e+00  3.3536595 0.69225281     P   0
## Year:Block!Year_79   2.601849e+00  4.1605472 0.62536224     P   0
## Year:Block!Year_80   8.148340e-01  1.5000504 0.54320444     P   0
## Year:Block!Year_81   1.028062e+01 16.8785574 0.60909369     P   0
## Year:Block!Year_82   3.782552e-07         NA         NA     B   0
## Year:Block!Year_83   1.311193e-01  0.8770122 0.14950686     P   0
## Year:Block!Year_84   3.149464e+00  4.8633243 0.64759494     P   0
## Year:Block!Year_85   3.782552e-07         NA         NA     B   0
## Year:Block!Year_86   5.980740e-06         NA         NA     B   0
## Year:Block!Year_87   4.368026e-05         NA         NA     B   0
## Year:Block!Year_88   2.563949e-01  3.5621790 0.07197699     P   0
## Year:Block:Wholeplot!Year_77      3.782552e-07 NA          NA     B   0
## Year:Block:Wholeplot!Year_79  7.705295e-01  1.2528742   0.61500946   P   0
## Year:Block:Wholeplot!Year_80  7.394376e-01  0.8696140   0.85030549   P   0
## Year:Block:Wholeplot!Year_81   7.413468e+00  5.7598117   1.28710245   P   0
## Year:Block:Wholeplot!Year_82   2.701891e-01  1.3808788   0.19566460   P   0
## Year:Block:Wholeplot!Year_83  1.617967e+00  1.6054174   1.00781705   P   0
## Year:Block:Wholeplot!Year_84   5.980740e-06   NA       NA          B   0
## Year:Block:Wholeplot!Year_85 5.980740e-06  NA       NA          B   0
```

```
## Year:Block:Wholeplot!Year_86      1.481352e-05        NA       NA              B    0
## Year:Block:Wholeplot!Year_87    5.832377e+01    38.6155750    1.51036903     P    0
## Year:Block:Wholeplot!Year_88     7.262868e+00   7.3523383 0.98783110   P    0
## Year:Block:NWholeplot:Subplot(R)   1.000000e+00      NA      NA      F    0
## Year:Block:Nwholeplot:Subplot!Year_77 7.466666e-01  0.2822136 2.64575043 P   0
## Year:Block:Nwholeplot:Subplot!Year_79 2.718222e+00  1.2156259 2.23606798   P   0
## Year:Block:Nwholeplot:Subplot!Year_80 1.368230e+00  0.6118913 2.23606798     P   0
## Year:Block:Nwholeplot:Subplot!Year_81 2.158255e+00  0.9652010 2.23606798     P   0
## Year:Block:Nwholeplot:Subplot!Year_82 4.794074e+00  2.1439751 2.23606798     P   0
## Year:Block:Nwholeplot:Subplot!Year_83 1.855457e+00  0.8297855 2.23606798     P   0
## Year:Block:Nwholeplot:Subplot!Year_84 4.328390e+00  1.6359796 2.64574820     P   0
## Year:Block:Nwholeplot:Subplot!Year_85 5.397298e+00  1.9708165 2.73861005     P   0
## Year:Block:Nwholeplot:Subplot!Year_86 1.417377e+01 5.1755335 2.73860958    P   0
## Year:Block:Nwholeplot:Subplot!Year_87 8.108572e+00  3.6262636 2.23606798   P   0
## Year:Block:Nwholeplot:Subplot!Year_88 8.893420e+00  3.9772582 2.23606798   P   0
vcnames <- rownames(summary(model5.asr)$varcomp)[1:22]
model5.asr <- setvarianceterms(model5.asr$call, terms = vcnames, bounds = "U",
                               initial=0.01, ignore.suffices = FALSE)

## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:33 2017
##           LogLik     Sigma2     DF      wall      cpu
##   1      -393.211      1.0     165 15:55:33     0.0 (18 restrained)
##   2      -367.175      1.0     165 15:55:33     0.0 (9 restrained)
##   3      -338.725      1.0     165 15:55:33     0.0 (7 restrained)
##   4      -303.755      1.0     165 15:55:33     0.0 (8 restrained)
##   5      -285.353      1.0     165 15:55:33     0.0 (7 restrained)

## Warning in asreml(fixed = Yield ~ Year * Rotation * Nitrogen, random =
~idh(Year):Block/Wholeplot, :
## Singularity in average information matrix

## 2 singularities in Average Information matrix
##   6      -278.943      1.0     165 15:55:33     0.0 (7 restrained)
##   7      -276.627      1.0     165 15:55:33     0.0 (6 restrained)
##   8      -275.985      1.0     165 15:55:33     0.0 (3 restrained)
##   9      -276.843      1.0     165 15:55:33     0.0 (2 restrained)
## 10      -277.981      1.0     165 15:55:33     0.0 (2 restrained)
## 11      -279.125      1.0     165 15:55:33     0.0 (2 restrained)
## 12      -280.273      1.0     165 15:55:33     0.0 (2 restrained)
## 13      -281.422      1.0     165 15:55:33     0.0 (2 restrained)

## Warning in asreml(fixed = Yield ~ Year * Rotation * Nitrogen, random =
~idh(Year):Block/Wholeplot, : Log-
## likelihood not converged

## Warning in asreml(fixed = Yield ~ Year * Rotation * Nitrogen, random =
~idh(Year):Block/Wholeplot, : Some
## components changed by more than 1% on the last iteration.

model5.asr <- update(model5.asr)

## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:33 2017
##         LogLik     Sigma2   DF   wall     cpu
##  1      -282.573     1 .0    165 15:55:33    0.0 (2 restrained)

## Warning in asreml(fixed = Yield ~ Year * Rotation * Nitrogen, random =
~idh(Year):Block/Wholeplot, :
## Singularity in average information matrix

## 1 singularities in Average Information matrix
##  2   -282.543                 1.0     165 15:55:33    0.0 (2 restrained)
##   3 -282.518                  1.0     165 15:55:33    0.0 (2 restrained)
##   4        -282.499           1.0     165 15:55:33    0.0 (2 restrained)
##   5        -282.487           1.0     165 15:55:33    0.0 (1 restrained)
##   6        -282.483           1.0     165 15:55:33    0.0 (1 restrained)
##   7        -282.483           1.0     165 15:55:33    0.0 (1 restrained)
summary(model5.asr)$varcomp
```

```
##                                       component     std.error   z.ratio     bound %ch
## Year:Block!Year_77                     2.344716e+00  3.3536596   0.69915147  U    0.0
## Year:Block!Year_79                     2.601849e+00  4.1605472   0.62536224  U    0.0
## Year:Block!Year_80                     8.148340e-01  1.5000504   0.54320444  U    0.0
## Year:Block!Year_81                     1.028062e+01 16.8785574   0.60909369  U    0.0
## Year:Block!Year_82                     1.074595e-01  0.4935380  -0.21773306  U    0.0
## Year:Block!Year_83                     1.311193e-01  0.8770122   0.14950686  U    0.0
## Year:Block!Year_84                     3.152837e+00  4.8662813   0.64789447  U    0.0
## Year:Block!Year_85                    -1.372373e-01  0.2891636  -0.47460095  U    0.0
## Year:Block!Year_86                    -3.118268e+06  NA          NA          S    0.0
## Year:Block!Year_87                    -3.176159e+00 16.4108486  -0.19354020  B    0.1
## Year:Block!Year_88                     2.563949e-01  3.5621790   0.07197699  U    0.0
## Year:Block:Wholeplot!Year_77          -1.619506e-01  0.1621616  -0.99869904  U    0.0
## Year:Block:Wholeplot!Year_79           7.705295e-01  1.2528742   0.61500946  U    0.0
## Year:Block:Wholeplot!Year_80           7.394376e-01  0.8696140   0.85030549  U    0.0
## Year:Block:Wholeplot!Year_81           7.413468e+00  5.7598117   1.28710245  U    0.0
## Year:Block:Wholeplot!Year_82           3.776488e-01  1.5691971   0.24066374  U    0.0
## Year:Block:Wholeplot!Year_83           1.617967e+00  1.6054174   1.00781705  U    0.0
## Year:Block:Wholeplot!Year_84          -2.360219e-02  1.1987019  -0.01968980  U    0.0
## Year:Block:Wholeplot!Year_85          -6.354705e-01  1.3623072  -0.46646637  U    0.0
## Year:Block:Wholeplot!Year_86          -9.913303e-01  3.8453095  -0.25780247  U    0.0
## Year:Block:Wholeplot!Year_87           6.149949e+01 45.4102563   1.35430844  U    0.0
## Year:Block:Wholeplot!Year_88           7.262868e+00  7.3523383   0.98783110  U    0.0
## Year:Block:NWholeplot:Subplot(R)       1.000000e+00  NA          NA          F    0.0
## Year:Block:NWholeplot:Subplot!Year_77  8.854815e-01  0.3959994   2.23606798  P    0.0
## Year:Block:NWholeplot:Subplot!Year_79  2.718222e+00  1.2156259   2.23606798  P    0.0
## Year:Block:NWholeplot:Subplot!Year_80  1.368230e+00  0.6118913   2.23606798  P    0.0
## Year:Block:NWholeplot:Subplot!Year_81  2.158255e+00  0.9652010   2.23606798  P    0.0
## Year:Block:NWholeplot:Subplot!Year_82  4.794074e+00  2.1439751   2.23606798  P    0.0
## Year:Block:NWholeplot:Subplot!Year_83  1.855457e+00  0.8297855   2.23606798  P    0.0
## Year:Block:NWholeplot:Subplot!Year_84  4.348626e+00  1.9447645   2.23606798  P    0.0
## Year:Block:NWholeplot:Subplot!Year_85  6.170008e+00  2.7593116   2.23606798  P    0.0
## Year:Block:NWholeplot:Subplot!Year_86  1.584683e+01  7.0869141   2.23606875  P    0.0
## Year:Block:NWholeplot:Subplot!Year_87  8.108572e+00  3.6262636   2.23606798  P    0.0
## Year:Block:NWholeplot:Subplot!Year_88  8.893420e+00  3.9772582   2.23606798  P    0.0
```

```
info.accumulate <- rbind(info.accumulate,
                         data.frame(Model = "Meta analysis with
different variance components",
                                    infoCriteria(model5.asr, bound.
exclusions = "F"),
                                    stringsAsFactors = FALSE))
## Warning in infoCriteria.asreml(model5.asr, bound.exclusions =
"F"): The following bound terms were discounted:
##  Year:Block:NWholeplot:Subplot(R)
Accumulated summary of REML information criteria
info.accumulate
##Model                                           DF NBound AIC      BIC      logREML
## 1 Split-plot                                    3  0     653.1797 662.4975 -323.5899
## 2 Split plot meta analysis                      13 1     614.3496 654.7269 -294.1748
## 3 Split-plot and EXP                            4  0     655.1797 667.6035 -323.5899
## 4 Split-plot meta analysis and EXP              14 1     616.0131 659.4963 -294.0065
## 5 Meta analysis with different variance components 33 1 630.9657 733.4619 -282.4829
```

Use split plot with different residual variance in each year

```
wald(model2.asr, denDF = "algebraic")
## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:33 2017
##          LogLik        Sigma2     DF     wall     cpu
##  1      -294.175          1.0    165 15:55:33     0.0
##  2      -294.175          1.0    165 15:55:33     0.0
##  3      -294.175          1.0    165 15:55:33     0.2
## $Wald
##
## Wald tests for fixed effects.
## Response: Yield
##
##                         Df denDF   F.inc       Pr
## (Intercept)              1  10.5 15300.0 0.000000
## Year                    10  10.4    85.4 0.000000
## Rotation                 4  27.0    26.7 0.000000
## Nitrogen                 2  58.2    55.3 0.000000
## Year:Rotation           40  22.7     3.7 0.000683
## Year:Nitrogen           20  48.7     1.3 0.241290
## Rotation:Nitrogen        8  58.2     1.9 0.072879
## Year:Rotation:Nitrogen 80  40.6     1.2 0.270253
##
## $stratumVariances
## NULL
```

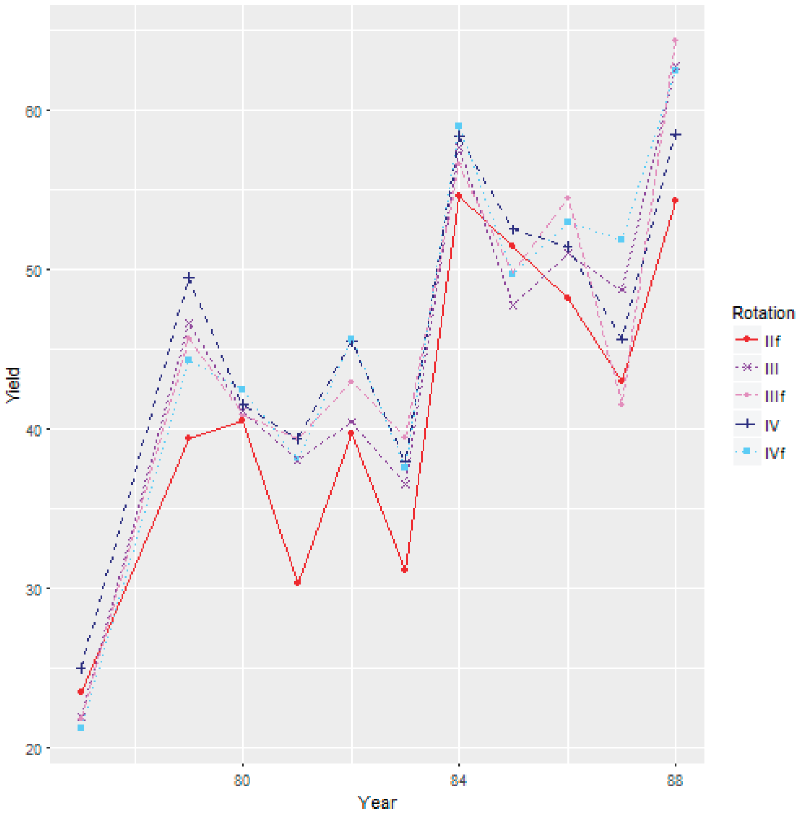
Drop unnecessary fixed terms

```
model2a.asr <- asreml(Yield ~ Year*Rotation*Nitrogen-Year:Rotation:Nitrogen,
                   random = ~ Year:Block/Wholeplot,
                   residual = ~ idh(Year):Block:Wholeplot:Subplot,
                       data = wmpotato)
## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:34 2017
##          LogLik        Sigma2     DF     wall     cpu
##  1      -454.257          1.0    245 15:55:34     0.0
##  2      -436.705          1.0    245 15:55:34     0.0
##  3      -432.022          1.0    245 15:55:34     0.0
##  4      -429.833          1.0    245 15:55:34     0.0
##  5      -428.217          1.0    245 15:55:34     0.0
##  6      -426.955          1.0    245 15:55:34     0.0
##  7      -426.438          1.0    245 15:55:34     0.0
##  8      -426.321          1.0    245 15:55:34     0.0
##  9      -426.301          1.0    245 15:55:34     0.0
## 10      -426.297          1.0    245 15:55:34     0.0
## 11      -426.296          1.0    245 15:55:34     0.0
## 12      -426.296          1.0    245 15:55:34     0.0
wald(model2a.asr, denDF = "algebraic")
## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:34 2017
##          LogLik        Sigma2     DF     wall     cpu
##  1      -426.296          1.0    245 15:55:34     0.0
##  2      -426.296          1.0    245 15:55:34     0.0
##  3      -426.296          1.0    245 15:55:34     0.0
## $Wald
##
## Wald tests for fixed effects.
## Response: Yield
##
##                    Df denDF   F.inc       Pr
## (Intercept)         1  10.7 14670.0 0.000000
## Year               10  10.7    79.8 0.000000
## Rotation            4  32.8    23.0 0.000000
## Nitrogen            2 128.4    41.9 0.000000
## Year:Rotation      40  25.3     2.9 0.002830
## Year:Nitrogen      20  88.3     1.4 0.166185
## Rotation:Nitrogen   8 128.4     1.3 0.252994
##
## $stratumVariances
## NULL
model2b.asr <- asreml(Yield ~ Year*Rotation*Nitrogen-Year:Rotation:Nitrogen
-
                                        Rotation:Nitrogen,
              random = ~ Year:Block/Wholeplot,
```

```
                  residual =   ~ idh(Year):Block:Nwholeplot:Subplot,
                          data = wmpotato)
## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:34 2017
##            LogLik        Sigma2     DF      wall     cpu
##  1       -458.206           1.0    253 15:55:34    0.0 (1 restrained)
##  2       -452.705           1.0    253 15:55:34    0.0
##  3       -440.278           1.0    253 15:55:34    0.0
##  4       -433.140           1.0    253 15:55:34    0.0
##  5       -429.272           1.0    253 15:55:34    0.0
##  6       -427.727           1.0    253 15:55:34    0.0
##  7       -427.310           1.0    253 15:55:34    0.0
##  8       -427.236           1.0    253 15:55:34    0.0
##  9       -427.226           1.0    253 15:55:34    0.0
## 10       -427.225           1.0    253 15:55:34    0.0
## 11        -427.224           1.0   253 15:55:34    0.0
wald(model2b.asr, denDF = "algebraic")
## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:34 2017
##            LogLik        Sigma2     DF      wall     cpu
##  1       -427.224           1.0    253 15:55:34    0.0
##  2       -427.224           1.0    253 15:55:34    0.0
##  3       -427.224           1.0    253 15:55:34    0.0
## $Wald
##
## Wald tests for fixed effects.
## Response: Yield
##
##                 Df denDF   F.inc       Pr
## (Intercept)      1  10.8 14460.0 0.000000
## Year            10  10.8    78.5 0.000000
## Rotation         4  34.0    22.4 0.000000
## Nitrogen         2 136.7    41.9 0.000000
## Year:Rotation   40  26.9     2.8 0.003045
## Year:Nitrogen   20  95.6     1.4 0.154287
##
## $stratumVariances
## NULL
model2c.asr <- asreml(Yield ~ Year*Rotation*Nitrogen-Year:Rotation:Nitrogen
-
                        Rotation:Nitrogen - Year:Nitrogen,
                      random = ~ Year:Block/Wholeplot,
                     residual =   ~ idh(Year):Block:Nwholeplot:Subplot,
                      data = wmpotato)
## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:34 2017
##            LogLik        Sigma2     DF      wall     cpu
##  1      -482.365          1.0     273 15:55:34    0.0 (1 restrained)
##  2       -473.437           1.0    273 15:55:34    0.0
##  3       -457.815           1.0    273 15:55:34    0.0
##  4       -449.143           1.0    273 15:55:34    0.0
##  5       -444.554           1.0    273 15:55:34    0.0
##  6       -443.052           1.0    273 15:55:34    0.0
##  7       -442.746           1.0    273 15:55:34    0.0
##  8       -442.702           1.0    273 15:55:34    0.0
##  9       -442.695           1.0    273 15:55:34    0.0
## 10       -442.694           1.0    273 15:55:34    0.0
## 11       -442.694           1.0    273 15:55:34    0.0
wald(model2c.asr, denDF = "algebraic")
## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:34 2017
##            LogLik        Sigma2     DF      wall     cpu
##  1       -442.694           1.0    273 15:55:34    0.0
##  2       -442.694           1.0    273 15:55:34    0.0
##  3       -442.694           1.0    273 15:55:34    0.0
## $Wald
##
## Wald tests for fixed effects.
## Response: Yield
##
##                 Df denDF   F.inc         Pr
## (Intercept)      1  10.8 14360.0 0.00000000
## Year            10  10.8    78.2 0.00000001
## Rotation         4  32.6    23.2 0.00000000
## Nitrogen         2 142.0    41.6 0.00000000
```

```
## Year:Rotation 40  27.0     2.9 0.00223137
##
## $stratumVariances
## NULL
Get predictions and plot
predict(model2c.asr, classify = "Nitrogen")$pvals
## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:34 2017
##          LogLik        Sigma2     DF     wall     cpu
##  1      -442.694          1.0    273 15:55:34    0.0
##  2      -442.694          1.0    273 15:55:34    0.0
##  3      -442.694          1.0    273 15:55:34    0.0
##
## Notes:
## - The predictions are obtained by averaging across the hypertable
##   calculated from model terms constructed solely from factors in
##   the averaging and classify sets.
## - Use ‘average’ to move ignored factors into the averaging set.
## - The simple averaging set: Year,Rotation
## - The ignored set: Block,Wholeplot
## ##
##   Nitrogen predicted.value std.error    status
## 1       N1        43.36754  0.402073 Estimable
## 2       N2        45.02538  0.402073 Estimable
## 3       N3        45.64905  0.402073 Estimable
predYR <- predict(model2c.asr, classify = "Year:Rotation")$pvals
## Model fitted using the sigma parameterization.
## ASReml4 Beta-release 4.0.0.9005 Mon Jan 23 15:55:34 2017
##          LogLik        Sigma2     DF     wall     cpu
##  1      -442.694          1.0    273 15:55:34    0.0
##  2      -442.694          1.0    273 15:55:34    0.0
##  3      -442.694          1.0    273 15:55:34    0.0
predYR
##
## Notes:
## - The predictions are obtained by averaging across the hypertable
##   calculated from model terms constructed solely from factors in
##   the averaging and classify sets.
## - Use 'average' to move ignored factors into the averaging set.
## - The simple averaging set: Nitrogen
## - The ignored set: Block,Wholeplot
##
## ##    Year Rotation predicted.value std.error    status
## 1     77      IIf        23.43704  1.485051 Estimable
## 2     77      III        21.91852  1.485051 Estimable
## 3     77     IIIf        21.87037  1.485051 Estimable
## 4     77       IV        25.01481  1.485051 Estimable
## 5     77      IVf        21.19630  1.485051 Estimable
## 6     79      IIf        39.44074  1.533421 Estimable
## 7     79      III        46.61852  1.533421 Estimable
## 8     79     IIIf        45.73704  1.533421 Estimable
## 9     79       IV        49.50370  1.533421 Estimable
## 10    79      IVf        44.35185  1.533421 Estimable
## 11    80      IIf        40.52593  1.448644 Estimable
## 12    80      III        41.13704  1.448644 Estimable
## 13    80     IIIf        40.89630  1.448644 Estimable
## 14    80       IV        41.54444  1.448644 Estimable
## 15    80      IVf        42.48148  1.448644 Estimable
## 16    81      IIf        30.37407  1.546351 Estimable
## 17    81      III        38.04444  1.546351 Estimable
## 18    81     IIIf        39.45185  1.546351 Estimable
## 19    81       IV        39.42963  1.546351 Estimable
## 20    81      IVf        38.13704  1.546351 Estimable
## 21    82      IIf        39.75185  1.594618 Estimable
## 22    82      III        40.48519  1.594618 Estimable
## 23    82     IIIf        42.94815  1.594618 Estimable
## 24    82       IV        45.53333  1.594618 Estimable
## 25    82      IVf        45.62963  1.594618 Estimable
## 26    83      IIf        31.14444  1.535587 Estimable
## 27    83      III        36.54815  1.535587 Estimable
## 28    83     IIIf        39.52963  1.535587 Estimable
## 29    83       IV        37.95556  1.535587 Estimable
## 30    83      IVf        37.61481  1.535587 Estimable
## 31    84      IIf        54.58889  1.634753 Estimable
## 32    84      III        57.67407  1.634753 Estimable
```

```
## 33   84      IIIf        56.63333  1.634753 Estimable
## 34   84        IV        58.40000  1.634753 Estimable
## 35   84       IVf        58.99259  1.634753 Estimable
## 36   85       IIf        51.48148  1.781401 Estimable
## 37   85       III        47.80000  1.781401 Estimable
## 38   85      IIIf        49.78519  1.781401 Estimable
## 39   85        IV        52.53704  1.781401 Estimable
## 40   85       IVf        49.75185  1.781401 Estimable
## 41   86       IIf        48.20926  2.132871 Estimable
## 42   86       III        51.02778  2.132871 Estimable
## 43   86      IIIf        54.57593  2.132871 Estimable
## 44   86        IV        51.48148  2.132871 Estimable
## 45   86       IVf        52.98704  2.132871 Estimable
## 46   87       IIf        42.98519  2.809408 Estimable
## 47   87       III        48.75556  2.809408 Estimable
## 48   87      IIIf        41.55556  2.809408 Estimable
## 49   87        IV        45.67407  2.809408 Estimable
## 50   87       IVf        51.85000  2.809408 Estimable
## 51   88       IIf        54.28333  1.917713 Estimable
## 52   88       III        62.76111  1.917713 Estimable
## 53   88      IIIf        64.41407  1.917713 Estimable
## 54   88        IV        58.52407  1.917713 Estimable
## 55   88       IVf        62.45556  1.917713 Estimable
cols <- c('red','darkviolet','violet','darkblue','lightskyblue')
ggplot(data = predYR,
       aes(x =Year, y=predicted.value, colour=Rotation, linetype=Rotation,
shape = Rotation)) +
  geom_point() + geom_line() + labs(y = "Yield") +
  scale_color_manual(values = cols) + scale_shape_manual(values = c(16,4,18,3,15))
```



Appendix Fig. 1. plot of chunk unnamed-chunk-11

Appendix 5. SAS code and output from the analysis of potato yields from the Westmaas experiment (supplemental provided by Kathleen Yeater)

```
/*Try Various Random Models*/
/*Split plot within years*/
title1 'Split-plot nested within years';
Proc Mixed data=wmpotato method=REML covtest;*method=REML is default;
class Year_ Block_ Wholeplot_ Rotation_ Nitrogen_;
model Yield = Year_|Rotation_|Nitrogen_ / ddfm=kr;
random Block_ Block_*Wholeplot_ / subject=Year_;
ods select covparms fitstatistics;
run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	*Z* Value	Pr > *Z*
Block_	Year_	1.4072	1.4027	1.00	0.1579
Block_*Wholeplot_	Year_	7.0791	1.8929	3.74	< 0.0001
Residual		5.1952	0.7005	7.42	< 0.0001

Fit Statistics	
-2 Res Log Likelihood	950.4
AIC (Smaller is Better)	956.4
AICC (Smaller is Better)	956.6
BIC (Smaller is Better)	957.6

```
/*Nested split-plot, different residual variance each year*/
title1 'Nested split plot with different residual variance in each
year';
Proc Mixed data=wmpotato method=REML covtest;
class Year_ Block_ Wholeplot_ Rotation_ Nitrogen_;
model Yield = Year_|Rotation_|Nitrogen_ / ddfm=kr;
random Block_ Block_*Wholeplot_ / subject=Year_;
repeated / group=Year_;
ods select covparms fitstatistics;
run;
```

Covariance Parameter Estimates						
Cov Parm	Subject	Group	Estimate	Standard Error	Z Value	Pr > Z
Block_	Year_		2.1173	1.1783	1.80	0.0362
Block_*Wholeplot_	Year_		0.9103	0.8170	1.11	0.1326
Residual		Year_ 77	0.8185	0.3435	2.38	0.0086
Residual		Year_ 79	2.6840	1.1544	2.32	0.0100
Residual		Year_ 80	1.3442	0.5864	2.29	0.0109
Residual		Year_ 81	5.2395	4.1243	1.27	0.1020
Residual		Year_ 82	4.5379	1.8554	2.45	0.0072
Residual		Year_ 83	1.9885	0.9469	2.10	0.0179
Residual		Year_ 84	4.0134	1.6360	2.45	0.0071
Residual		Year_ 85	5.4145	2.1357	2.54	0.0056
Residual		Year_ 86	14.2836	5.4923	2.60	0.0047
Residual		Year_ 87	58.2196	23.1443	2.52	0.0059
Residual		Year_ 88	12.8988	5.5662	2.32	0.0102

Fit Statistics	
-2 Res Log Likelihood	891.6
AIC (Smaller is Better)	917.6
AICC (Smaller is Better)	920.0
BIC (Smaller is Better)	922.8

```
/*Nested split plot with AR1 structure over years*/
title1 'Nested split-plot and AR1';
title2 ' ';
Proc Mixed data=wmpotato method=REML covtest;
class Year_ Block_ Wholeplot_ Rotation_ Nitrogen_;
model Yield = Year_|Rotation_|Nitrogen_ / ddfm=kr;
random Block_ Block_*Wholeplot_ / subject=Year_ ;
repeated / subject=Year_ type=AR(1);
ods select covparms fitstatistics;
run;
```

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
Block_	Year_	1.3777	1.3946	0.99	0.1616
Block_*Wholeplot_	Year_	6.7744	2.0190	3.36	0.0004
AR(1)	Year_	0.06903	0.1568	0.44	0.6598
Residual		5.4732	1.0339	5.29	< 0.0001

Fit Statistics	
-2 Res Log Likelihood	950.2
AIC (Smaller is Better)	958.2
AICC (Smaller is Better)	958.5
BIC (Smaller is Better)	959.8

```
/*Nested split plot with different residual variance in each year AND
AR1 structure*/
title1 'Nested split-plot with heterogeneous AR1 structure in each
year';
Proc Mixed data=wmpotato method=REML covtest;
class Year_ Block_ Wholeplot_ Rotation_ Nitrogen_;
model Yield = Year_|Rotation_|Nitrogen_ / ddfm=kr;
random Block_ Block_*Wholeplot_ / subject=Year_;
```

```
repeated / group=Year_ type=ARH(1);
ods select covparms fitstatistics;
run;
```

Covariance Parameter Estimates						
Cov Parm	Subject	Group	Estimate	Standard Error	*Z* Value	Pr *Z*
Block_	Year_		2.1173	1.1783	1.80	0.0362
Block_*Wholeplot_	Year_		0.9101	0.8167	1.11	0.1326
Var(1)		Year_ 77	0.8185	0.3435	2.38	0.0086
ARH(1)		Year_ 77	0	.	.	.
Var(1)		Year_ 79	2.6840	1.1544	2.32	0.0100
ARH(1)		Year_ 79	0	.	.	.
Var(1)		Year_ 80	1.3442	0.5863	2.29	0.0109
ARH(1)		Year_ 80	0	.	.	.
Var(1)		Year_ 81	5.2402	4.1239	1.27	0.1019
ARH(1)		Year_ 81	0	.	.	.
Var(1)		Year_ 82	4.5380	1.8553	2.45	0.0072
ARH(1)		Year_ 82	0	.	.	.
Var(1)		Year_ 83	1.9885	0.9469	2.10	0.0179
ARH(1)		Year_ 83	0	.	.	.
Var(1)		Year_ 84	4.0134	1.6360	2.45	0.0071
ARH(1)		Year_ 84	0	.	.	.
Var(1)		Year_ 85	5.4144	2.1356	2.54	0.0056
ARH(1)		Year_ 85	0	.	.	.
Var(1)		Year_ 86	14.2837	5.4923	2.60	0.0047
ARH(1)		Year_ 86	0	.	.	.
Var(1)		Year_ 87	58.2206	23.1442	2.52	0.0059
ARH(1)		Year_ 87	0	.	.	.
Var(1)		Year_ 88	12.8992	5.5662	2.32	0.0102
ARH(1)		Year_ 88	0	.	.	.

Fit Statistics	
-2 Res Log Likelihood	891.6
AIC (Smaller is Better)	939.6
AICC (Smaller is Better)	948.2
BIC (Smaller is Better)	949.1

```
/*Nested split-plot with different residual variances and variance
components*/
/*Note: This model is CPU-intensive, prepare for a longer than usual
run time*/
title1 'Nested split-plot with variance components in each year';
Proc Mixed data=wmpotato method=REML covtest;
class Year_ Block_ Wholeplot_ Rotation_ Nitrogen_;
model Yield = Year_|Rotation_|Nitrogen_ / ddfm=kr;
random Block_ Block_*Wholeplot_ / group=Year_ type=VC;
repeated / group=Year_ type=VC;
ods select covparms fitstatistics;
run;
```

Fit Statistics	
-2 Res Log Likelihood	854.8
AIC (Smaller is Better)	904.8
AICC (Smaller is Better)	914.2
BIC (Smaller is Better)	872.2

Covariance Parameter Estimates					
Cov Parm	Group	Estimate	Standard Error	*Z* Value	Pr > *Z*
Block_	Year_ 77	2.3216	3.3537	0.69	0.2444
Block_	Year_ 79	2.6018	4.1605	0.63	0.2659
Block_	Year_ 80	0.8148	1.5001	0.54	0.2935
Block_	Year_ 81	10.2806	16.8786	0.61	0.2712
Block_	Year_ 82	0	.	.	.

Covariance Parameter Estimates					
Cov Parm	Group	Estimate	Standard Error	*Z* Value	Pr > *Z*
Block_	Year_ 83	0.1311	0.8770	0.15	0.4406
Block_	Year_ 84	3.1495	4.8633	0.65	0.2586
Block_	Year_ 85	0	.	.	.
Block_	Year_ 86	0	.	.	.
Block_	Year_ 87	0	.	.	.
Block_	Year_ 88	0.2564	3.5622	0.07	0.4713
Block_*Wholeplot_	Year_ 77	0	.	.	.
Block_*Wholeplot_	Year_ 79	0.7705	1.2529	0.62	0.2693
Block_*Wholeplot_	Year_ 80	0.7394	0.8696	0.85	0.1976
Block_*Wholeplot_	Year_ 81	7.4135	5.7598	1.29	0.0990
Block_*Wholeplot_	Year_ 82	0.2702	1.3809	0.20	0.4224
Block_*Wholeplot_	Year_ 83	1.6180	1.6054	1.01	0.1568
Block_*Wholeplot_	Year_ 84	0	.	.	.
Block_*Wholeplot_	Year_ 85	0	.	.	.
Block_*Wholeplot_	Year_ 86	0	.	.	.
Block_*Wholeplot_	Year_ 87	58.3238	38.6156	1.51	0.0655
Block_*Wholeplot_	Year_ 88	7.2629	7.3523	0.99	0.1616
Residual	Year_ 77	0.7464	0.2820	2.65	0.0041
Residual	Year_ 79	2.7182	1.2156	2.24	0.0127
Residual	Year_ 80	1.3682	0.6119	2.24	0.0127
Residual	Year_ 81	2.1583	0.9652	2.24	0.0127
Residual	Year_ 82	4.7941	2.1440	2.24	0.0127
Residual	Year_ 83	1.8555	0.8298	2.24	0.0127
Residual	Year_ 84	4.3284	1.6360	2.65	0.0041
Residual	Year_ 85	5.3971	1.9707	2.74	0.0031
Residual	Year_ 86	14.1737	5.1755	2.74	0.0031
Residual	Year_ 87	8.1086	3.6263	2.24	0.0127
Residual	Year_ 88	8.8934	3.9773	2.24	0.0127

```
/*Nested split-plot, different residual variance each year*/
/*Tests for Fixed Effects*/
title1 'Nested split plot with different residual variance in each
year';
title2 'Tests for Fixed Effects';
Proc Mixed data=wmpotato method=REML covtest;
class Year_ Block_ Wholeplot_ Rotation_ Nitrogen_;
model Yield = Year_|Rotation_|Nitrogen_ / ddfm=kr chisq;
random intercept Block_ Block_*Wholeplot_ / subject=Year_;
repeated / group=Year_;
ods select tests3;
run;
```

Type 3 Tests of Fixed Effects						
Effect	Num DF	Den DF	Chi-Square	*F Value*	Pr > ChiSq	Pr > *F*
Year_	10	9.98	851.71	85.52	< 0.0001	< 0.0001
Rotation_	4	44.4	71.00	17.75	< 0.0001	< 0.0001
Year_*Rotation_	40	6.51	154.96	4.49	< 0.0001	0.0269
Nitrogen_	2	35.8	21.19	10.59	< 0.0001	0.0002
Year_*Nitrogen_	20	32.1	29.16	1.24	0.0847	0.2841
Rotation_*Nitrogen_	8	35.8	6.30	0.79	0.6137	0.6167
Year_*Rotati*Nitroge	80	17.5	108.44	1.22	0.0189	0.3278

```
title1 'Nested split plot with different residual variance in each
year';
title2 'Drop 3-way Fixed term';
Proc Mixed data=wmpotato method=REML covtest;
class Year_ Block_ Wholeplot_ Rotation_ Nitrogen_;
model Yield = Year_|Rotation_|Nitrogen_ @2 / ddfm=kr chisq;
```

```
random intercept Block_ Block_*Wholeplot_ / subject=Year_;
repeated / group=Year_;
ods select tests3;
run;
```

Type 3 Tests of Fixed Effects						
Effect	Num DF	Den DF	Chi-Square	*F* Value	Pr > ChiSq	Pr > *F*
Year_	10	10.5	796.19	79.81	< 0.0001	< 0.0001
Rotation_	4	52.9	71.73	17.93	< 0.0001	< 0.0001
Year_*Rotation_	40	14.9	117.03	3.02	< 0.0001	0.0122
Nitrogen_	2	68.4	27.46	13.73	< 0.0001	< 0.0001
Year_*Nitrogen_	20	79.5	29.67	1.35	0.0754	0.1747
Rotation_*Nitrogen_	8	126	8.40	1.05	0.3954	0.4025

```
title1 'Nested split plot with different residual variance in each year';
title2 '3-way Fixed term and Rotation*Nitrogen Removed';
Proc Mixed data=wmpotato method=REML covtest;
class Year_ Block_ Wholeplot_ Rotation_ Nitrogen_;
model Yield = Year_ Rotation_ Nitrogen_ Year_*Rotation_
Year_*Nitrogen_ / ddfm=kr chisq;
random intercept Block_ Block_*Wholeplot_ / subject=Year_;
repeated / group=Year_;
ods select tests3;
run;
```

Type 3 Tests of Fixed Effects						
Effect	Num DF	Den DF	Chi-Square	*F* Value	Pr > ChiSq	Pr > *F*
Year_	10	10.8	783.24	78.50	< 0.0001	< 0.0001
Rotation_	4	57.2	69.67	17.42	< 0.0001	< 0.0001
Nitrogen_	2	70.7	27.45	13.72	< 0.0001	< 0.0001
Year_*Rotation_	40	18.4	112.77	2.87	< 0.0001	0.0089
Year_*Nitrogen_	20	89.4	29.82	1.37	0.0729	0.1588

```
title1 'Nested split plot with different residual variance in each
year';
title2 '3-way Fixed term, Rotation*Nitrogen, and Year*Nitrogen
Removed';
Proc Mixed data=wmpotato method=REML covtest;
class Year_ Block_ Wholeplot_ Rotation_ Nitrogen_;
model Yield = Year_ Rotation_ Nitrogen_ Year_*Rotation_ / ddfm=kr
chisq;
random intercept Block_ Block_*Wholeplot_ / subject=Year_;
repeated / group=Year_;
ods select tests3;
run;
```

Type 3 Tests of Fixed Effects						
Effect	Num DF	Den DF	Chi-Square	*F* Value	Pr > ChiSq	Pr > *F*
Year_	10	10.8	780.44	78.22	< 0.0001	< 0.0001
Rotation_	4	59.7	70.33	17.58	< 0.0001	< 0.0001
Nitrogen_	2	140	70.72	35.36	< 0.0001	< 0.0001
Year_*Rotation_	40	17.4	117.47	2.99	< 0.0001	0.0081

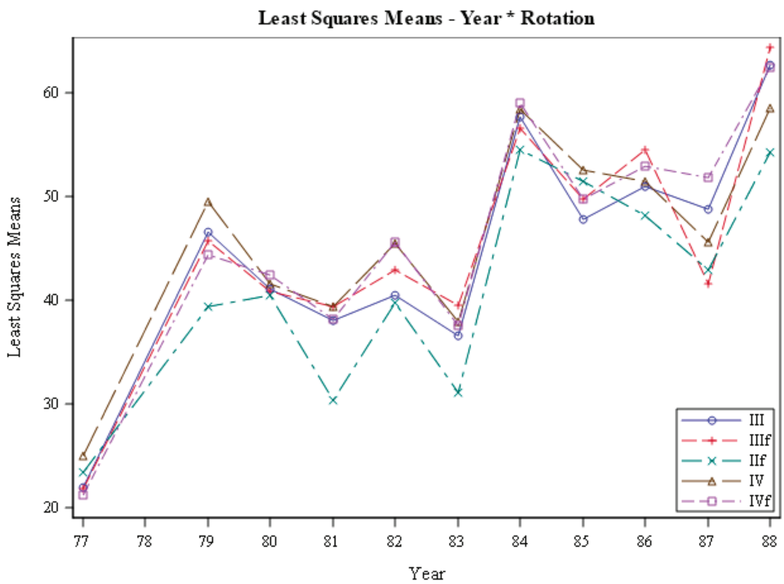
```
/*Get predictions and plot*/
/*Predicted means for Nitrogen*/
/*Predicted means for Year*Rotation*/
title1 'Predicted Means for Nitrogen';
title2 'Predicted Means for Year*Rotation';
Proc Mixed data=wmpotato method=REML covtest;
class Year_ Block_ Wholeplot_ Rotation_ Nitrogen_;
model Yield = Year_ Rotation_ Nitrogen_ Year_*Rotation_ / ddfm=kr
chisq;
```

```
random intercept Block_ Block_*Wholeplot_ / subject=Year_;
repeated / group=Year_;
lsmeans Nitrogen_ Year_*Rotation_;
ods select lsmeans;
run;
```

Least Squares Means

Effect	Rotation_	Nitrogen_	Year_	Estimate	Standard Error	DF	*t* Value	Pr > |*t*|
Nitrogen_		N1		43.3676	0.4069	16.2	106.57	< 0.0001
Nitrogen_		N2		45.0254	0.4069	16.2	110.65	< 0.0001
Nitrogen_		N3		45.6490	0.4069	16.2	112.18	< 0.0001
Year_*Rotation_	III		77	21.9185	1.4849	18.3	14.76	< 0.0001
Year_*Rotation_	IIIf		77	21.8704	1.4849	18.3	14.73	< 0.0001
Year_*Rotation_	IIf		77	23.4370	1.4849	18.3	15.78	< 0.0001
Year_*Rotation_	IV		77	25.0148	1.4849	18.3	16.85	< 0.0001
Year_*Rotation_	IVf		77	21.1963	1.4849	18.3	14.27	< 0.0001
Year_*Rotation_	III		79	46.6185	1.5333	20.6	30.40	< 0.0001
Year_*Rotation_	IIIf		79	45.7370	1.5333	20.6	29.83	< 0.0001
Year_*Rotation_	IIf		79	39.4407	1.5333	20.6	25.72	< 0.0001
Year_*Rotation_	IV		79	49.5037	1.5333	20.6	32.29	< 0.0001
Year_*Rotation_	IVf		79	44.3519	1.5333	20.6	28.93	< 0.0001
Year_*Rotation_	III		80	41.1370	1.4485	16.9	28.40	< 0.0001
Year_*Rotation_	IIIf		80	40.8963	1.4485	16.9	28.23	< 0.0001
Year_*Rotation_	IIf		80	40.5259	1.4485	16.9	27.98	< 0.0001
Year_*Rotation_	IV		80	41.5444	1.4485	16.9	28.68	< 0.0001
Year_*Rotation_	IVf		80	42.4815	1.4485	16.9	29.33	< 0.0001
Year_*Rotation_	III		81	38.0444	1.5463	23.7	24.60	< 0.0001
Year_*Rotation_	IIIf		81	39.4519	1.5463	23.7	25.51	< 0.0001
Year_*Rotation_	IIf		81	30.3741	1.5463	23.7	19.64	< 0.0001
Year_*Rotation_	IV		81	39.4296	1.5463	23.7	25.50	< 0.0001
Year_*Rotation_	IVf		81	38.1370	1.5463	23.7	24.66	< 0.0001
Year_*Rotation_	III		82	40.4852	1.5945	23.2	25.39	< 0.0001
Year_*Rotation_	IIIf		82	42.9481	1.5945	23.2	26.94	< 0.0001
Year_*Rotation_	IIf		82	39.7519	1.5945	23.2	24.93	< 0.0001
Year_*Rotation_	IV		82	45.5333	1.5945	23.2	28.56	< 0.0001
Year_*Rotation_	IVf		82	45.6296	1.5945	23.2	28.62	< 0.0001
Year_*Rotation_	III		83	36.5481	1.5354	20.9	23.80	< 0.0001
Year_*Rotation_	IIIf		83	39.5296	1.5354	20.9	25.74	< 0.0001
Year_*Rotation_	IIf		83	31.1444	1.5354	20.9	20.28	< 0.0001
Year_*Rotation_	IV		83	37.9556	1.5354	20.9	24.72	< 0.0001
Year_*Rotation_	IVf		83	37.6148	1.5354	20.9	24.50	< 0.0001
Year_*Rotation_	III		84	57.6741	1.6346	24.9	35.28	< 0.0001
Year_*Rotation_	IIIf		84	56.6333	1.6346	24.9	34.65	< 0.0001
Year_*Rotation_	IIf		84	54.5889	1.6346	24.9	33.40	< 0.0001
Year_*Rotation_	IV		84	58.4000	1.6346	24.9	35.73	< 0.0001
Year_*Rotation_	IVf		84	58.9926	1.6346	24.9	36.09	< 0.0001
Year_*Rotation_	III		85	47.8000	1.7813	29.8	26.84	< 0.0001
Year_*Rotation_	IIIf		85	49.7852	1.7813	29.8	27.95	< 0.0001
Year_*Rotation_	IIf		85	51.4815	1.7813	29.8	28.90	< 0.0001
Year_*Rotation_	IV		85	52.5370	1.7813	29.8	29.49	< 0.0001
Year_*Rotation_	IVf		85	49.7519	1.7813	29.8	27.93	< 0.0001
Year_*Rotation_	III		86	51.0278	2.1328	36.1	23.93	< 0.0001
Year_*Rotation_	IIIf		86	54.5759	2.1328	36.1	25.59	< 0.0001
Year_*Rotation_	IIf		86	48.2093	2.1328	36.1	22.60	< 0.0001
Year_*Rotation_	IV		86	51.4815	2.1328	36.1	24.14	< 0.0001
Year_*Rotation_	IVf		86	52.9870	2.1328	36.1	24.84	< 0.0001
Year_*Rotation_	III		87	48.7556	2.8095	31.5	17.35	< 0.0001
Year_*Rotation_	IIIf		87	41.5556	2.8095	31.5	14.79	< 0.0001
Year_*Rotation_	IIf		87	42.9852	2.8095	31.5	15.30	< 0.0001
Year_*Rotation_	IV		87	45.6741	2.8095	31.5	16.26	< 0.0001
Year_*Rotation_	IVf		87	51.8500	2.8095	31.5	18.46	< 0.0001
Year_*Rotation_	III		88	62.7611	1.9176	35.1	32.73	< 0.0001
Year_*Rotation_	IIIf		88	64.4141	1.9176	35.1	33.59	< 0.0001

Least Squares Means								
Effect	Rotation_	Nitrogen_	Year_	Estimate	Standard Error	DF	*t* Value	Pr > |*t*|
Year_*Rotation_	IIf		88	54.2833	1.9176	35.1	28.31	< 0.0001
Year_*Rotation_	IV		88	58.5241	1.9176	35.1	30.52	< 0.0001
Year_*Rotation_	IVf		88	62.4556	1.9176	35.1	32.57	< 0.0001




```
/*Plot of year(x) predicted value(y) rotation (group) */

data YR;
set lsmeans;
if Effect = 'Nitrogen_' then delete;
run;

proc template;
define statgraph sgdesign;
dynamic _YEAR_ _ESTIMATE _ROTATION_;
begingraph;
   entrytitle halign=center 'Least Squares Means - Year * Rotation';
   layout lattice / rowdatarange=data columndatarange=data
rowgutter=10 columngutter=10;
      layout overlay / xaxisopts=( label=('Year') linearopts=(
tickvaluesequence=( start=77.0 end=88.0 increment=1.0))) yaxisopts=(
label=('Least Squares Means'));
         seriesplot x=_YEAR_ y=_ESTIMATE / group=_ROTATION_
name='series' display=(markers) clusterwidth=0.5 connectorder=xaxis
grouporder=data;
         discretelegend 'series' / opaque=false border=true
halign=right valign=bottom displayclipped=true across=1 order=rowmajor
location=inside;
      endlayout;
   endlayout;
endgraph;
end;
run;

proc sgrender data=YR template=sgdesign;
dynamic _YEAR_="'YEAR_'n" _ESTIMATE="ESTIMATE" _
ROTATION_="'ROTATION_'n";
run;
```

Appendix 6. Answers to Review Questions

1. Why would you do a long-term rotation experiment instead of several single-year experiments?

   Rotation experiments allow you to study differences between sequences of treatments that are applied over several years. They also allow you to study how the effect of a treatment develops over more than a year.

2. What is the difference between a short-term and a long-term rotation experiment?

   In a short-term rotation experiment the sequences run through one simultaneous cycle, to compare the sequences in the final year. Long-term rotation experiment run through several cycles, and involve analyses of data from more than a year.

3. Why might it be a problem if you ran the rotations over only one series of years?

   The comparisons between the rotations will depend on the specific properties of those years, which may favor one rotation over the others. There is less of a risk of this happening if you run the experiments over more than one series of years.

4. How might you include auxiliary treatments, in addition to the rotation treatments?

   The simplest way to do this is to split the plots into subplots, to form a split-plot design in each year, with the auxiliary treatment factor(s) as the split-plot factor(s).

5. Why might the analysis be more complicated than the analysis of a single-year experiment?

   The results will be recorded from several different years, and these may show different amounts of random variation. The same plot may be observed in several years and, unless these observations are well separated, the results may show a nonuniform correlation structure where the correlations between these observations decline with increasing distance in time.

6. What are the advantages of REML compared to ordinary analysis of variance, and how would exploit these in your analysis?

   REML allows different residual variances to be estimated for the years during the combined analysis of data from several years. It also allows you to fit models to describe the correlations between observations at different times on the same plot.

7. What statistics can you use to decide on the random model?

   You can use the differences between the deviances of two models, if one is a generalization of the other (i.e. if it contains all the random parameters of that model) together with some additional ones. This can be treated can be treated as a chi-square statistic with number of degrees of freedom equal to the number of additional parameters. Otherwise you can use Akaike or Schwarz Bayesian information criteria. The best model is the one with the smallest value of the chosen criterion.

8. How would you assess the fixed terms?

   The standard way to do this is to examine their Wald statistics. These would have exact chi-square distributions if the variance parameters were known but, as those must be estimated, the statistics are only asymptotically distributed as chi-square. In practical situations they are biased (i.e. they tend to give too many significant results). Alternatively, your statistical software may be able to estimate the number of residual degrees of freedom relevant to each term, so that F statistics can be used instead. These should not be subject to the biases of the chi-square statistics.

CHAPTER 12: SPATIAL STATISTICS OF FIELD EXPERIMENTS

Juan Burgueño

SAS programs for Examples are provided in the electronic supplement.

Answer to review questions

1. Explain with your own words what means "spatial variability analysis"

   In field experimental designs it is related with modelling residual error by considering the spatial distribution of the plots in the field.

2. What are the different models that can be used to model the residual spatial variability?

   Among others, neighbor models, moving average, autoregressive models in row and columns, splines.

3. Why modelling the residual spatial variability usually has larger precision than a standard analysis in which spatial variability is not modeling?

   Because the experimental design is not able to capture all the variability in the field. It is not able to capture variability generated during the experimentation and it is not abel to capture small-scale variability.

4. Is it possible to model spatial variability of an experiment with five treatments in two replicates? Justify your answer.

   It is difficult since there is not so much information, there is a few number of row and columns in the field to adjust most of the model used to model spatial variability.

5. If you are analyzing an experimental design; and you want to perform spatial analysis, are there any changes in the assumptions? If so, mention them.

   Yes, all spatial analysis models assume some degree of relatedness between residuals error compared with the assumption of independence used in standard analyses.

6. What are the advantages of using spatial analysis?

   With spatial analysis it is possible to capture more noise and extract more information about the treatments. Spatial analysis usually is more precise and it adjusts the means of the treatments by the position in the field.

CHAPTER 13: AUGMENTED DESIGNS-EXPERIMENTAL DESIGNS IN WHICH ALL TREATMENTS ARE NOT REPLICATED

Juan Burgueño, José Crossa, Francisco Rodríguez, and Kathleen M. Yeater

Appendix A1. Complete results for Example 1 obtained with the codes in the text. SAS editor file (.sas) provided in electronic supplement.

(P1-f)
The GLIMMIX Procedure

Model Information	
Data Set	WORK.A
Response Variable	y
Response Distribution	Gaussian
Link Function	Identity
Variance Function	Default
Variance Matrix	Diagonal
Estimation Technique	Restricted Maximum Likelihood
Degrees of Freedom Method	Residual

Class Level Information		
Class	Levels	Values
t	8	1 2 3 4 5 6 7 8

Number of Observations Read	12
Number of Observations Used	12

Dimensions	
Covariance Parameters	1
Columns in X	9
Columns in Z	0
Subjects (Blocks in V)	1
Max Obs per Subject	12

Optimization Information	
Optimization Technique	None
Parameters	9
Lower Boundaries	1
Upper Boundaries	0
Fixed Effects	Not Profiled

Fit Statistics	
-2 Res Log Likelihood	16.94
AIC (smaller is better)	34.94
AICC (smaller is better)	214.94
BIC (smaller is better)	29.41
CAIC (smaller is better)	38.41
HQIC (smaller is better)	22.82
Pearson Chi-Square	9.33
Pearson Chi-Square / DF	2.33

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	*F* Value	Pr > *F*
t	7	4	16.06	0.0088

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > |t|
Ch	-5.3333	1.2472	4	-4.28	0.0129
UnT	-4.0000	2.1602	4	-1.85	0.1377
Ch-UnT	-5.6667	1.7638	4	-3.21	0.0325

t Least Squares Means					
t	Estimate	Standard Error	DF	t Value	Pr > |t|
1	22.3333	0.8819	4	25.32	< 0.0001
2	27.6667	0.8819	4	31.37	< 0.0001
3	28.0000	1.5275	4	18.33	< 0.0001
4	32.0000	1.5275	4	20.95	< 0.0001
5	27.0000	1.5275	4	17.68	< 0.0001
6	38.0000	1.5275	4	24.88	< 0.0001
7	35.0000	1.5275	4	22.91	< 0.0001
8	28.0000	1.5275	4	18.33	< 0.0001

Appendix 1. Complete results for Example 1 obtained with the codes in the text.

(P2-f)

The GLIMMIX Procedure

Model Information	
Data Set	WORK.A
Response Variable	y
Response Distribution	Gaussian
Link Function	Identity
Variance Function	Default
Variance Matrix	Diagonal
Estimation Technique	Restricted Maximum Likelihood
Degrees of Freedom Method	Residual

Class Level Information		
Class	Levels	Values
t	8	1 2 3 4 5 6 7 8
d1	3	1 2 3

Number of Observations Read	12
Number of Observations Used	12

Dimensions	
Covariance Parameters	1
Columns in X	12
Columns in Z	0
Subjects (Blocks in V)	1
Max Obs per Subject	12

Optimization Information	
Optimization Technique	None
Parameters	9
Lower Boundaries	1
Upper Boundaries	0
Fixed Effects	Not Profiled

Fit Statistics	
-2 Res Log Likelihood	16.94
AIC (smaller is better)	34.94
AICC (smaller is better)	214.94

Fit Statistics	
BIC (smaller is better)	29.41
CAIC (smaller is better)	38.41
HQIC (smaller is better)	22.82
Pearson Chi-Square	9.33
Pearson Chi-Square / DF	2.33

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	*F* Value	Pr > *F*
d1	2	4	34.93	0.0029
t(d1)	5	4	8.51	0.0295

Estimates					
Label	Estimate	Standard Error	DF	*t* Value	Pr > |*t*|
Ch	-5.3333	1.2472	4	-4.28	0.0129
UnT	-4.0000	2.1602	4	-1.85	0.1377
Ch-UnT	-5.6667	1.7638	4	-3.21	0.0325

d1 Least Squares Means					
d1	Estimate	Standard Error	DF	t Value	Pr > |*t*|
1	22.3333	0.8819	4	25.32	< 0.0001
2	27.6667	0.8819	4	31.37	< 0.0001
3	31.3333	0.6236	4	50.25	< 0.0001

t(d1) Least Squares Means						
t	d1	Estimate	Standard Error	DF	t Value	Pr > |*t*|
1	1	22.3333	0.8819	4	25.32	< 0.0001
2	2	27.6667	0.8819	4	31.37	< 0.0001
3	3	28.0000	1.5275	4	18.33	< 0.0001
4	3	32.0000	1.5275	4	20.95	< 0.0001
5	3	27.0000	1.5275	4	17.68	< 0.0001
6	3	38.0000	1.5275	4	24.88	< 0.0001
7	3	35.0000	1.5275	4	22.91	< 0.0001
8	3	28.0000	1.5275	4	18.33	< 0.0001

Appendix A1. Complete results for Example 1 obtained with the codes in the text.

(P3-f)

The GLIMMIX Procedure

Model Information	
Data Set	WORK.A
Response Variable	y
Response Distribution	Gaussian
Link Function	Identity
Variance Function	Default
Variance Matrix	Diagonal
Estimation Technique	Restricted Maximum Likelihood
Degrees of Freedom Method	Residual

Class Level Information		
Class	Levels	Values
t	8	1 2 3 4 5 6 7 8
d2	2	1 2

Number of Observations Read	12
Number of Observations Used	12

Dimensions	
Covariance Parameters	1
Columns in X	11
Columns in Z	0

Dimensions	
Subjects (Blocks in V)	1
Max Obs per Subject	12

Optimization Information	
Optimization Technique	None
Parameters	9
Lower Boundaries	1
Upper Boundaries	0
Fixed Effects	Not Profiled

Fit Statistics	
-2 Res Log Likelihood	16.94
AIC (smaller is better)	34.94
AICC (smaller is better)	214.94
BIC (smaller is better)	29.41
CAIC (smaller is better)	38.41
HQIC (smaller is better)	22.82
Pearson Chi-Square	9.33
Pearson Chi-Square / DF	2.33

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	*F* Value	Pr > *F*
d2	1	4	51.57	0.0020
t(d2)	6	4	10.14	0.0210

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
Ch	-5.3333	1.2472	4	-4.28	0.0129
UnT	-4.0000	2.1602	4	-1.85	0.1377
Ch-UnT	-5.6667	1.7638	4	-3.21	0.0325

d2 Least Squares Means

d2	Estimate	Standard Error	DF	*t* Value	Pr > |*t*|
1	25.0000	0.6236	4	40.09	< 0.0001
2	31.3333	0.6236	4	50.25	< 0.0001

t(d2) Least Squares Means

t	d2	Estimate	Standard Error	DF	*t* Value	Pr > |*t*|
1	1	22.3333	0.8819	4	25.32	< 0.0001
2	1	27.6667	0.8819	4	31.37	< 0.0001
3	2	28.0000	1.5275	4	18.33	< 0.0001
4	2	32.0000	1.5275	4	20.95	< 0.0001
5	2	27.0000	1.5275	4	17.68	< 0.0001
6	2	38.0000	1.5275	4	24.88	< 0.0001
7	2	35.0000	1.5275	4	22.91	< 0.0001
8	2	28.0000	1.5275	4	18.33	< 0.0001

Tests of Effect Slices for t(d2) Sliced By d2

d2	Num DF	Den DF	*F* Value	Pr > *F*
1	1	4	18.29	0.0129
2	5	4	8.51	0.0295

Appendix A1. Complete results for Example 1 obtained with the codes in the text.

(P2-r)

The GLIMMIX Procedure

Model Information	
Data Set	WORK.A
Response Variable	Y
Response Distribution	Gaussian
Link Function	Identity

Model Information	
Variance Function	Default
Variance Matrix	Not blocked
Estimation Technique	Restricted Maximum Likelihood
Degrees of Freedom Method	Containment

Class Level Information		
Class	Levels	Values
t	8	1 2 3 4 5 6 7 8
d1	3	1 2 3

Number of Observations Read	12
Number of Observations Used	12

Dimensions	
G-side Cov. Parameters	1
R-side Cov. Parameters	1
Columns in X	4
Columns in Z	8
Subjects (Blocks in V)	1
Max Obs per Subject	12

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	1
Lower Boundaries	1
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History					
Iteration	Restarts	Evaluations	Objective Function	Change	Max Gradient
0	**0**	4	47.864285513	.	2.05E-15

Convergence criterion (ABSGCONV=0.00001) satisfied.

Fit Statistics	
-2 Res Log Likelihood	47.86
AIC (smaller is better)	51.86
AICC (smaller is better)	53.86
BIC (smaller is better)	52.02
CAIC (smaller is better)	54.02
HQIC (smaller is better)	50.79
Generalized Chi-Square	21.00
Gener. Chi-Square / DF	2.33

Covariance Parameter Estimates		
Cov Parm	Estimate	Standard Error
t(d1)	17.5333	12.6726
Residual	2.3333	1.6499

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	*F* Value	Pr > *F*
d1	2	5	2.00	0.2304

Solution for Random Effects							
Effect	t	d1	Estimate	Std Err Pred	DF	t Value	Pr > |t|
t(d1)	1	1	-626E-16	4.1873	4	-0.00	1.0000
t(d1)	2	2	-4E-14	4.1873	4	-0.00	1.0000
t(d1)	3	3	-2.9418	2.1537	4	-1.37	0.2437
t(d1)	4	3	0.5884	2.1537	4	0.27	0.7982
t(d1)	5	3	-3.8244	2.1537	4	-1.78	0.1504
t(d1)	6	3	5.8837	2.1537	4	2.73	0.0523
t(d1)	7	3	3.2360	2.1537	4	1.50	0.2074

Solution for Random Effects							
Effect	t	d1	Estimate	Std Err Pred	DF	t Value	Pr > |t|
t(d1)	8	3	-2.9418	2.1537	4	-1.37	0.2437

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > |t|
Ch	-5.3333	6.0516	5	-0.88	0.4185
UnT	-3.5302	2.0294	4	-1.74	0.1569
Ch-UnT	-6.0582	4.5184	5	-1.34	0.2377

d1 Least Squares Means					
d1	Estimate	Standard Error	DF	t Value	Pr > |t|
1	22.3333	4.2791	5	5.22	0.0034
2	27.6667	4.2791	5	6.47	0.0013
3	31.3333	1.8196	5	17.22	< 0.0001

Tests of Covariance Parameters Based on the Restricted Likelihood					
Label	DF	-2 Res Log Like	ChiSq	Pr > ChiSq	Note
t(d1=3)	1	51.9494	4.09	0.0216	MI

MI: P-value based on a mixture of chi-squares.

Appendix A1. Complete results for Example 1 obtained with the codes in the text.

(P3-r)

The GLIMMIX Procedure

Model Information	
Data Set	WORK.A
Response Variable	y
Response Distribution	Gaussian
Link Function	Identity
Variance Function	Default
Variance Matrix	Not blocked
Estimation Technique	Restricted Maximum Likelihood
Degrees of Freedom Method	Containment

Class Level Information		
Class	Levels	Values
t	8	1 2 3 4 5 6 7 8
d2	2	1 2

Number of Observations Read	12
Number of Observations Used	12

Dimensions	
G-side Cov. Parameters	2
R-side Cov. Parameters	1
Columns in X	3
Columns in Z	16
Subjects (Blocks in V)	1
Max Obs per Subject	12

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2
Upper Boundaries	0
Fixed Effects	Profiled
Residual Variance	Profiled
Starting From	Data

Iteration History					
Iteration	Restarts	Evaluations	Objective Function	Change	Max Gradient
0	**0**	4	54.050115446	.	5.27E-16

Convergence criterion (ABSGCONV=0.00001) satisfied.

Fit Statistics	
-2 Res Log Likelihood	54.05
AIC (smaller is better)	60.05
AICC (smaller is better)	64.05
BIC (smaller is better)	60.29
CAIC (smaller is better)	63.29
HQIC (smaller is better)	58.44
Generalized Chi-Square	23.33
Gener. Chi-Square / DF	2.33

Covariance Parameter Estimates			
Cov Parm	Group	Estimate	Standard Error
t(d2)	d2 1	13.4444	20.1208
t(d2)	d2 2	17.5333	12.6726
Residual		2.3333	1.6499

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	*F* Value	Pr > *F*
d2	1	6	3.85	0.0975

Solution for Random Effects								
Effect	t	d2	Group	Estimate	Std Err Pred	DF	t Value	Pr > |t|
t(d2)	1	1	d2 1	-2.5208	2.6627	4	-0.95	0.3974
t(d2)	2	1	d2 1	2.5208	2.6627	4	0.95	0.3974
t(d2)	3	2	d2 1	0	3.6667	4	0.00	1.0000
t(d2)	4	2	d2 1	0	3.6667	4	0.00	1.0000
t(d2)	5	2	d2 1	0	3.6667	4	0.00	1.0000
t(d2)	6	2	d2 1	0	3.6667	4	0.00	1.0000
t(d2)	7	2	d2 1	0	3.6667	4	0.00	1.0000
t(d2)	8	2	d2 1	0	3.6667	4	0.00	1.0000
t(d2)	1	1	d2 2	0	4.1873	4	0.00	1.0000
t(d2)	2	1	d2 2	0	4.1873	4	0.00	1.0000
t(d2)	3	2	d2 2	-2.9418	2.1537	4	-1.37	0.2437
t(d2)	4	2	d2 2	0.5884	2.1537	4	0.27	0.7982
t(d2)	5	2	d2 2	-3.8244	2.1537	4	-1.78	0.1504
t(d2)	6	2	d2 2	5.8837	2.1537	4	2.73	0.0523
t(d2)	7	2	d2 2	3.2360	2.1537	4	1.50	0.2074
t(d2)	8	2	d2 2	-2.9418	2.1537	4	-1.37	0.2437

Estimates					
Label	Estimate	Standard Error	DF	*t* Value	Pr > |*t*|
Ch	-5.0417	6.0446	4	-0.83	0.4512
UnT	3.5302	5.5684	4	0.63	0.5605
Ch-UnT	-5.9123	5.8172	6	-1.02	0.3487

d2 Least Squares Means					
d2	Estimate	Standard Error	DF	t Value	Pr > |*t*|
1	25.0000	2.6667	6	9.37	< 0.0001
2	31.3333	1.8196	6	17.22	< 0.0001

Tests of Covariance Parameters Based on the Restricted Likelihood					
Label	DF	-2 Res Log Like	ChiSq	Pr > ChiSq	Note
T(d2=1)	1	58.6165	4.57	0.0163	MI
T(d2=2)	1	58.1353	4.09	0.0216	MI

MI: P-value based on a mixture of chi-squares.

CHAPTER 14: MULTIVARIATE METHODS FOR AGRICULTURAL RESEARCH

Kathleen M. Yeater and Maria B. Villamil

Review Questions (True or False)

[*Hint*: Your T/F answers can be located using information in Table 1.]

1. We can explore relationships among variables in a data set with MR, CART, or even with cluster analysis among other tools. (T)

2. The main difference between PCA and FA is that the PCA tries to uncover theoretical constructs underlying the data set. (F)

3. SEM and CART are examples of flexible modern techniques that can integrate categorical and continuous variables. (T)

4. MANOVA, PCA, and DA are examples of techniques that require multivariate normality and homogeneity of variances. (T)

5. When the goal of the MA is to predict group membership, then DA, LR, or CART are NOT good choices. (F)

6. The desired result of applying MANOVA to a data set is to create linear combinations of variables that maximize our group differences. (F)

7. No previous knowledge of group membership is required when we explore the data set with CA. (T)

8. If the dependent variable in the data set only takes two values, a LR is the most appropriate technique to describe differences among the two possible outcomes. (T)

9. If your data set does not meet the requirements of multivariate normality you cannot apply any of the MA techniques. (F)

10. You can use CCPA to explore the distribution of several plant species on different environments characterized by topography, moisture level, aspect, etc. (T)

Exercises:

For the following exercises we will be working with the iris data set available from R by typing:

```
> library(datasets)
> iris
```

Once the data set is available on your workspace,

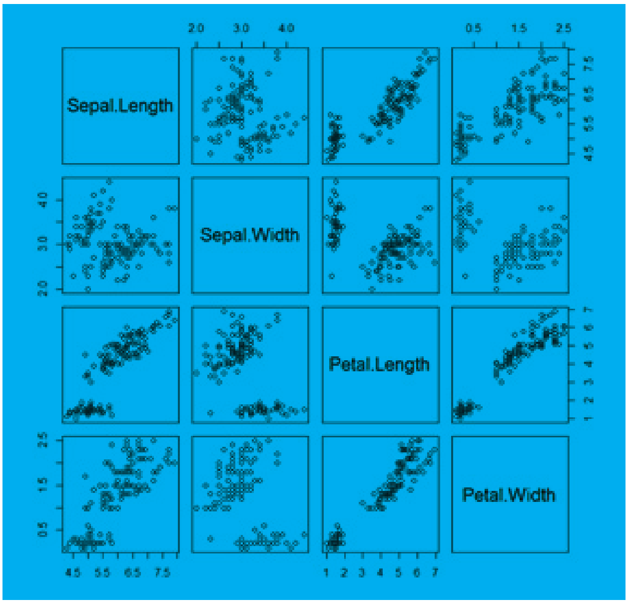
1) Investigate the structure of the data set and create a scatterplot matrix of the variables

`Sepal.Length, Sepal.Width, Petal.Length, Petal.Width.`

What can you infer from these results?

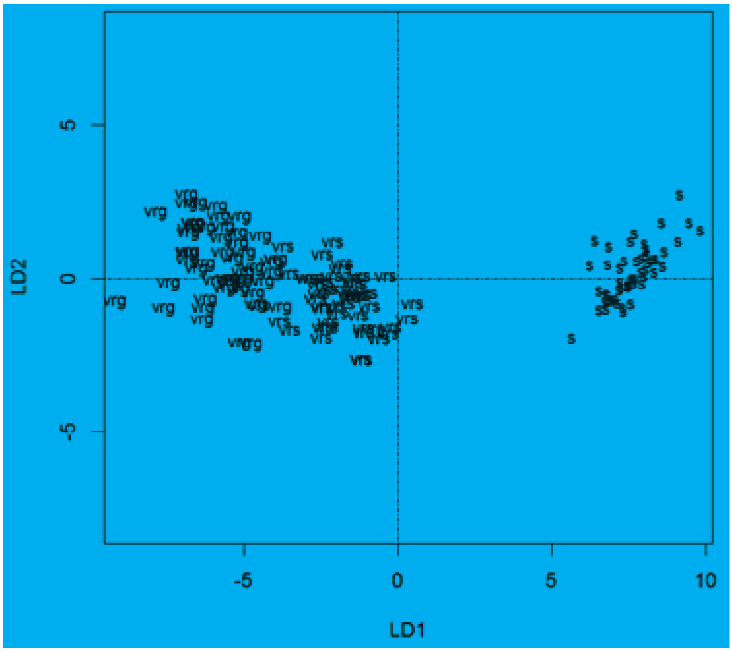
```
> str(iris)
'data.frame':   150 obs. of    5 variables:
```

```
$ Sepal.Length: num   5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
$ Sepal.Width : num   3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
$ Petal.Length: num   1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
$ Petal.Width : num   0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
$ Species      : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1
1 1 1 1 1 1 ...
> pairs(iris[,1:4])
```



2) Explore the possibility of successfully separating the Iris species based on the sepal and petal variables listed. Hint: You will need the MASS library.

```
> library(MASS)
> iris.lda<-lda(iris [,-5], iris[,5])
> iris.lda
Call:
lda(iris[, -5], iris[, 5])
Prior probabilities of groups:
    setosa versicolor  virginica
 0.3333333  0.3333333  0.3333333

Group means:
           Sepal.Length Sepal.Width Petal.Length Petal.Width
setosa            5.006       3.428        1.462       0.246
versicolor        5.936       2.770        4.260       1.326
virginica         6.588       2.974        5.552       2.026

Coefficients of linear discriminants:
                    LD1         LD2
Sepal.Length  0.8293776  0.02410215
Sepal.Width   1.5344731  2.16452123
Petal.Length -2.2012117 -0.93192121
Petal.Width  -2.8104603  2.83918785

Proportion of trace:
   LD1    LD2
0.9912 0.0088
> plot(iris.lda, abbrev=TRUE, cex=0.8) # Saved as Exercise lda plot.
tiff
> abline(h=0, v=0, lty=4)
```



3) Calculate the error rates associated with the classification of each of the Iris species and provide an interpretation of your results.

```
> iris.ld<-predict(iris.lda)
> iris.ldcv<-lda(iris [,-5], iris [,5], CV=T)
> table(true=iris$Species, pred=iris.ldcv$class)
            pred
true         setosa versicolor virginica
  setosa         50          0         0
  versicolor      0         48         2
  virginica       0          1        49
```

Essential website information

R: http://cran.r-project.org/web/views/Multivariate.html

SAS: http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm

CHAPTER 15: NONLINEAR REGRESSION MODELS AND APPLICATIONS

Fernando Miguez, Sotirios Archontoulis, and Hamze Dokoohaki

Answer To Exercises

Multiple choice correct answers

1. e)
2. d)
3. c)
4. b)
5. b)

Exercise 1

1.a. The simple exponential model is

$$Y = Y_0 \exp(-k)$$

The first partial derivative with respect to k is still

$$\frac{\partial Y}{\partial k} = Y_0 \exp(-kx) \times -x$$

Remember that the derivative of $f(x) = \exp(x)$ is also $\exp(x)$. The second derivative will not be equal to zero, therefore this is a function with a nonlinear parameter k.

1.b. In the model

$$y = \beta_0 + \beta_1 \exp(-\frac{x}{\theta})$$

The partial derivative with respect to the first parameter (β_0) is

$$\frac{\partial y}{\partial \beta_0} = 0$$

Thus, this is a linear parameter. The first partial derivative with respect to the second parameter () is

$$\frac{\partial y}{\partial \beta_1} = \exp\left(-\frac{x}{\theta}\right)$$

The second partial derivative is

$$\frac{\partial y}{\partial^2 \beta_1} = 0$$

Thus, this is a linear parameter. The partial derivative with respect to the parameter q is

$$\frac{\partial y}{\partial \theta} = x \times \frac{[\beta_1 \exp(-\frac{x}{\theta})]}{\theta^2}$$

Thus the second derivative will not be zero and this is not a linear parameter.

Exercise 2

1.a. Virtually any model which has an additive parameter such as

$$y = \beta_0 + \ldots$$

Will have the intercept model only as a subset. One less trivial example would be the exponential decay. Setting k=0 results in a model with a single parameter (Y_0). Another is the Michaelis–Menten equation in the following form

$$y = \frac{ax}{b+x}$$

where setting b = 0 results in a $y = a$

2.b.

It is tricky to show strictly that something cannot happen. A simple example could be

$$y = \exp(-kx)$$

In this case we do not have the simple intercept model as a subset of the full model. When k = 0, y = 1. As k increases, then y gets closer to zero. Again this is not an intercept model. Another example would be the one parameter logistic (Eq. 3.10, Table 3).

CHAPTER 16: ANALYSIS OF NON-GAUSSIAN DATA

Walter W. Stroup

Answers to Review Questions. Data sets provided as .csv file in electronic supplement.

1. a. Read "plot(block)|variety" as "plot within block after accounting for variety."

   b. Before applying treatments there are 2 plots and hence 1 df in each of 10 blocks. Therefore 10 df for plot(block). Accounting for variety removes the df for variety(in this case one. Therefore "plot(block)|variety" has $10-1=9$ df.
2. False. Leaving block*variety out of the RANDOM statement will result in overdispersion, a common form of poorly specified model.
3. False. Unless you include PLOT in the data set, the term PLOT(BLOCK*VARIETY) will be unintelligible to SAS. If you do include it, the algorithm GLIMMIX uses to determine denominator DF may not work properly, so you should check the listing. In any event, BLOCK*VARIETY uniquely identifies "plot(block)|variety" and avoids unintended consequences.
4. a. Y/N

   b. FALSE (!!!!)

   c. Make absolutely sure you have *both* Y *and* N in the data set for *every* experimental unit!
5. False. If you answered "true" Reread section on notation conventions for mixed models.
6. False.

   delete IRRIG from CLASS statement
7. cumulative logit
8. (b) leaf shape is a "nominal" multinomial variable (i.e. no obvious ranking or categories)
9. False. Science should drive statistics, not vice-versa. If science calls for unequal spacing, statistics can deal with it.
10. No
11. False. Using RANDOM...RESIDUAL will make this impossible.
12. True
13. False
14. False. (correct answer is Poisson)
15. True
16. True
17. True
18. False. If the distribution is binomial, The data scale estimate of the LSMEAN is the probability. (or proportion if the distribution is Beta)
19. True
20. False. Negative binomial is used for *count* data, not proportions. Use Beta for continuous proportion data.